

# Discovery of iridoid cyclase completes the iridoid pathway in asterids

Received: 12 June 2025

Accepted: 3 September 2025

Published online: 3 October 2025

Check for updates

Maite Colinas<sup>1</sup>✉, Chloée Tymen<sup>1</sup>, Joshua C. Wood<sup>2</sup>, Anja David<sup>1</sup>, Jens Wurlitzer<sup>1</sup>, Clara Morweiser<sup>1</sup>, Klaus Gase<sup>1</sup>, Ryan M. Alam<sup>1</sup>, Gabriel R. Titchiner<sup>1</sup>, John P. Hamilton<sup>2,3</sup>, Sarah Heinicke<sup>1</sup>, Ron P. Dirks<sup>4</sup>, Adriana A. Lopes<sup>5</sup>, Lorenzo Caputi<sup>1</sup>, C. Robin Buell<sup>2,6,7</sup>✉ & Sarah E. O'Connor<sup>1</sup>✉

Iridoids are specialized monoterpenes ancestral to asterid flowering plants<sup>1,2</sup> that play key roles in defence and are also essential precursors for pharmacologically important alkaloids<sup>3,4</sup>. The biosynthesis of all iridoids involves the cyclization of the reactive biosynthetic intermediate 8-oxocitronellyl enol. Here, using a variety of approaches including single-nuclei sequencing, we report the discovery of iridoid cyclases from a phylogenetically broad sample of asterid species that synthesize iridoids. We show that these enzymes catalyse formation of *7S-cis-trans* and *7R-cis-cis* nepetalactol, the two major iridoid stereoisomers found in plants. Our work uncovers a key missing step in the otherwise well-characterized early iridoid biosynthesis pathway in asterids. This discovery unlocks the possibility to generate previously inaccessible iridoid stereoisomers, which will enable metabolic engineering for the sustainable production of valuable iridoid and iridoid-derived compounds.

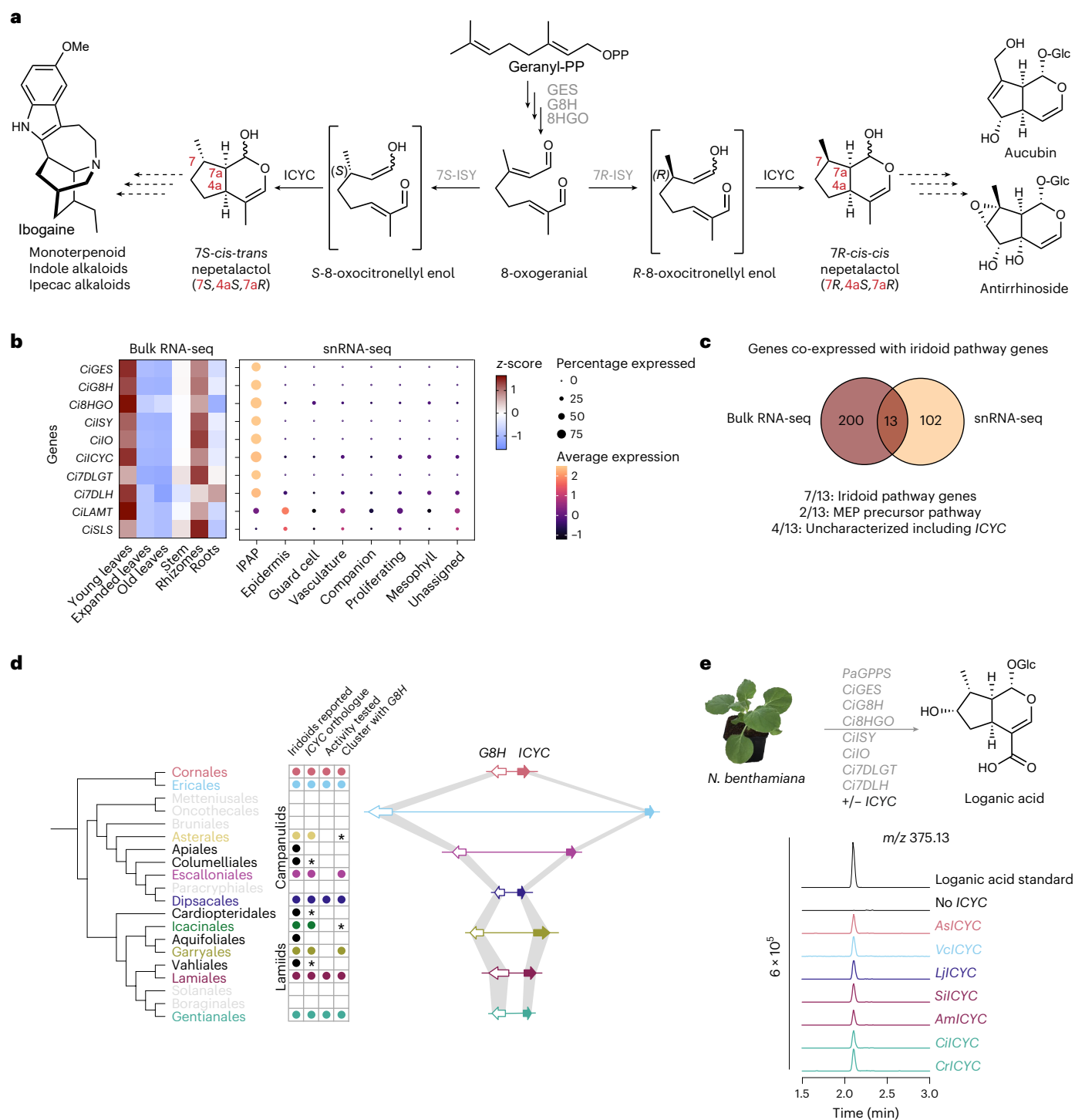
Iridoids are bicyclic monoterpenes widespread among asterid plants<sup>1,2</sup>. Iridoids play important roles in plant defence; volatile iridoids are used by plants to repel or attract insects, whereas glycosylated forms serve as feeding deterrents<sup>3–7</sup>. From a pharmacological perspective, iridoids possess promising anti-inflammatory activity<sup>8</sup> and, moreover, serve as precursors for medicinally important monoterpene indole and ipecac alkaloids, which include anti-cancer (camptothecin and vinblastine), anti-malarial (quinine), putative anti-addiction (ibogaine) and emetic (emetine) agents (recently reviewed by ref. 9).

Nepetalactol, which is the simplest iridoid and the common intermediate for all ~1,000 known iridoids, is biosynthesized from geranyl-diphosphate (GPP). GPP is subjected to dephosphorylation by geraniol synthase (GES), hydroxylation by geraniol 8-hydroxylase (G8H) and oxidation by 8-hydroxygeraniol oxidase (8HGO) to yield 8-oxo-geranial<sup>10–12</sup> (Fig. 1a and Supplementary Fig. 1). The short-chain

dehydrogenase iridoid synthase (ISY) catalyses a 1,4 reduction of 8-oxo-geranial to form 8-oxocitronellyl enol, which then cyclizes to form *7S/R*, *4aS* and *7aR*-nepetalactol, along with a number of side products<sup>13–15</sup>. Biosynthetic steps downstream of nepetalactol have been characterized in several plants, including *Catharanthus roseus* and *Camptotheca acuminata* (Supplementary Fig. 1)<sup>12,16–19</sup>.

Of all possible nepetalactol stereoisomers, *7S-cis-trans* (*7S*, *4aS*, *7aR*) and *7R-cis-cis* (*7R*, *4aS*, *7aR*) nepetalactol are most commonly observed in nature (Fig. 1a)<sup>1</sup>. *7S-cis-trans* nepetalactol-derived ('route I') iridoids, which are also precursors for alkaloid biosynthesis, are found among diverse asterid orders, whereas *7R-cis-cis* nepetalactol-derived ('route II') iridoids primarily occur in Lamiales families<sup>1,20,21</sup>. The stereo-configuration at the C-7 position is set by a lineage-specific stereoselective ISY that generates either *S*- or *R*-8-oxocitronellyl enol (catalysed by *7S*-ISY and *7R*-ISY, respectively)<sup>13,15</sup>. Although *7S-cis-trans* nepetalactol

<sup>1</sup>Department of Natural Product Biosynthesis, Max Planck Institute for Chemical Ecology, Jena, Germany. <sup>2</sup>Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA. <sup>3</sup>Department of Crop and Soil Sciences, University of Georgia, Athens, GA, USA. <sup>4</sup>Future Genomics Technologies, Leiden, The Netherlands. <sup>5</sup>Universidade de Ribeirão Preto, Departamento de Biotecnologia, Ribeirão Preto, Brazil. <sup>6</sup>Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, USA. <sup>7</sup>The Plant Center, University of Georgia, Athens, GA, USA. ✉e-mail: [mmartinez@ice.mpg.de](mailto:mmartinez@ice.mpg.de); [Robin.Buell@uga.edu](mailto:Robin.Buell@uga.edu); [occonnor@ice.mpg.de](mailto:occonnor@ice.mpg.de)



**Fig. 1 | Identification of ICYC. a**, Scheme showing iridoid biosynthesis including previously characterized iridoid pathway genes (grey). The 8-oxogeranial intermediate can be converted to either 7*S*-*cis*-*trans* nepetalactol or 7*R*-*cis*-*cis* nepetalactol by previously characterized species-specific stereoselective ISYs together with the newly identified ICYCs. For a complete scheme of the secoiridoid pathway including all intermediates, see Supplementary Fig. 1. **b**, *C. ipeacacuanha* bulk tissue RNA-seq and snRNA-seq show tight co-expression of orthologues of previously characterized secoiridoid pathway genes from *C. roseus* (Supplementary Fig. 3a) and the newly identified *CiICYC*. Complete tissue-specific data are shown in Supplementary Fig. 3b. Cell clusters were grouped into cell types that were determined using marker genes (see Supplementary Fig. 4 for complete analysis of snRNA-seq dataset). **c**, Venn diagram showing overlay of co-expressed genes from both datasets. See Extended Data Fig. 1a for a detailed list of genes. MEP, 2-C-methyl-D-erythritol 4-phosphate. **d**, Identification of *ICYC* orthologues

in various iridoid-producing orders. The tree is based on the latest angiosperm phylogeny<sup>35</sup>. Circles depict reported presence of iridoids within at least one species of the respective order. An asterisk denotes that no sequencing data were publicly available from a reported iridoid-producing species from the respective clade, and thus *ICYC* presence could not be determined. Right: a scheme depicting the synteny of the *ICYC* and *G8H* gene cluster. The synteny is shown for representative species of each order: *A. salviifolium* (Cornales), *V. corymbosum* (Ericales), *Escallonia rubra* (Escalloniales), *L. japonica* (Dipsacales), *Eucommia ulmoides* (Garryales), *A. majus* (Lamiales) and *C. ipeacacuanha* (Gentianales)<sup>100–104</sup>. **e**, Representative *ICYC* orthologues from different orders all reconstituted the iridoid pathway up to loganic acid in *N. benthamiana*. *Ci*, *C. ipeacacuanha*; *IO*, iridoid oxidase; *7DLH*, 7-deoxyloganic acid hydroxylase; *LAMT*, loganic acid methyltransferase; *SLS*, secologanin synthase; *As*, *A. salviifolium*; *Vc*, *V. corymbosum*; *Lj*, *L. japonica*; *Si*, *S. indicum*; *Am*, *A. majus*; *Ci*, *C. ipeacacuanha*; *Cr*, *C. roseus*.

is observed in low yields when 7S-*ISY* is incubated with 8-oxogeranial, it has long been hypothesized that an additional enzyme assists the cyclization of 3S- or 3R-8-oxocitronellyl enol to 7S-*cis-trans* and 7R-*cis-cis* nepetalactol for the following reasons: (a) spontaneous cyclization occurs only in vitro and almost exclusively yields the *cis-trans* configuration, thus failing to account for the existence of the 7R-*cis-cis* configuration when 7R-*ISY* is used with 8-oxogeranial<sup>15</sup>; (b) the known biosynthetic enzymes (GES, G8H, 8HGO, ISY) are not sufficient to reconstitute nepetalactol biosynthesis in *Nicotiana benthamiana*<sup>12,22</sup>; and (c) recent work on *Nepeta*, a Lamiaceae genus that independently evolved iridoid biosynthesis, identified *Nepeta*-specific proteins that assist formation of the 7S-*cis-trans* nepetalactol scaffold from S-8-oxocitronellyl enol<sup>14,23,24</sup>. Indeed, the inclusion of a *Nepeta* Major Latex Protein Like (MLPL) enabled the successful pathway reconstitution of 7S-iridoids and downstream alkaloids in yeast and *N. benthamiana*<sup>22,25</sup>.

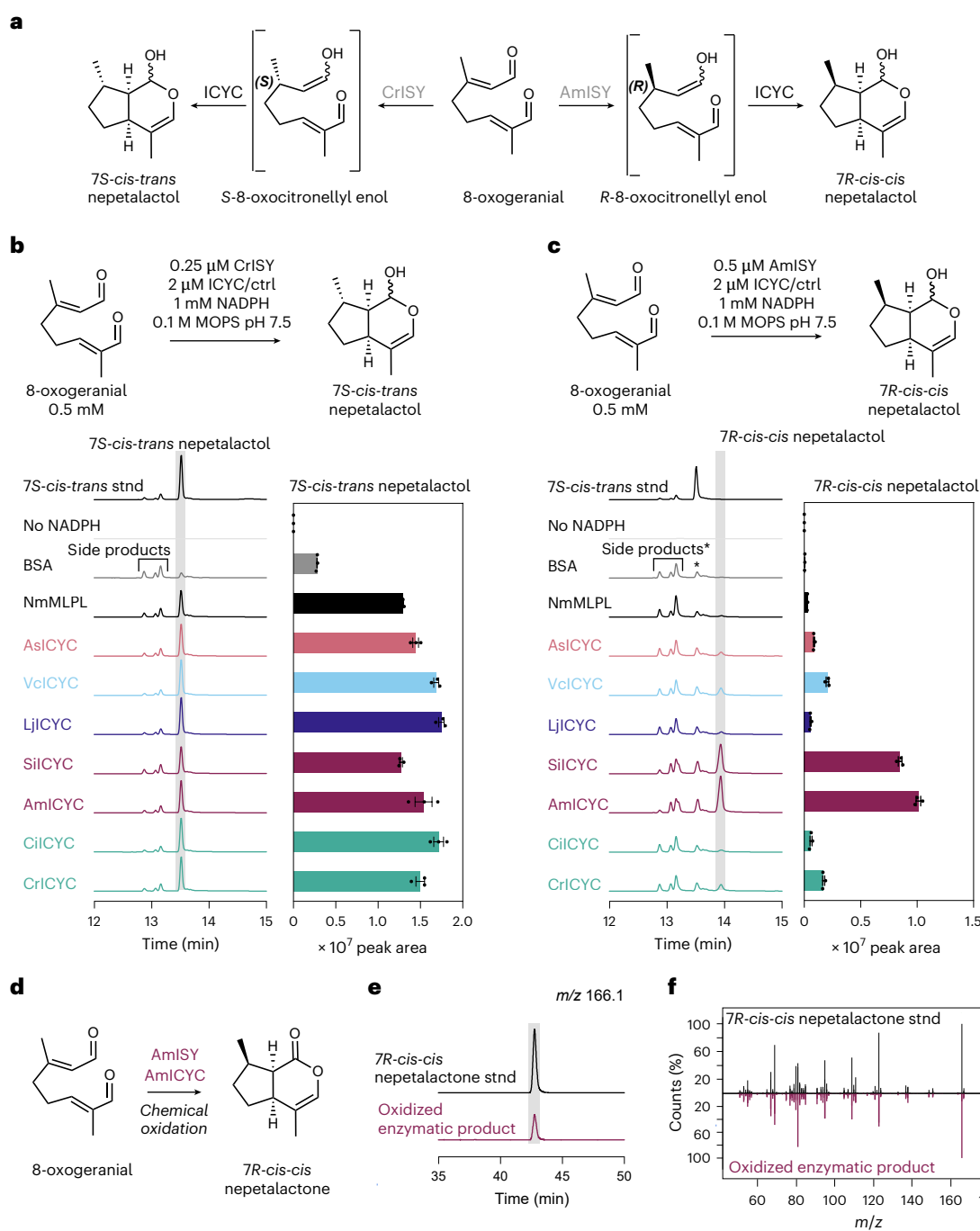
Identifying the asterid iridoid cyclase(s) (ICYCs) responsible for formation of 7S-*cis-trans* and 7R-*cis-cis* nepetalactol has been a long-standing challenge. The cyclization of this enol species to form the bicyclic scaffold that characterizes nepetalactol is an unusual chemical transformation that could be catalysed by any number of protein scaffolds. To obtain a refined pool of gene candidates for this biosynthetic step, we generated de novo genome assemblies and high-resolution expression data of two evolutionarily distant members of the asterid clade<sup>26</sup>, *Alangium salviifolium* (Cornales) and *Carapichea ipecacuanha* (Gentianales) along with single-nuclei transcriptomics of *C. ipecacuanha* (Supplementary Fig. 2–5 and Supplementary Tables 1–7). We identified highly conserved orthologues of *C. roseus* secoiridoid pathway genes in both species, confirming that iridoid biosynthesis is ancestral to asterids (Supplementary Fig. 3a). We next performed co-expression analysis using tissue-specific RNA-sequencing (RNA-seq) data and found iridoid pathway genes to be tightly co-expressed in *C. ipecacuanha* young leaves and rhizomes and in *A. salviifolium* roots (Fig. 1b and Supplementary Fig. 3b,c). In addition, we constructed gene co-expression networks on the cell clusters obtained from the single-nuclei RNA-seq (snRNA-seq) data of *C. ipecacuanha* young leaves (Fig. 1b, Supplementary Figs. 4 and 5 and Supplementary Dataset 1). Iridoid pathway orthologues up to 7-deoxyloganin acid hydroxylase (*ZDLH*) are tightly co-expressed in a cell cluster corresponding to the internal phloem associated parenchyma (IPAP) cells, a cell type that has been previously shown to harbour early and intermediate iridoid biosynthesis steps in *C. roseus*<sup>27</sup>; later iridoid pathway gene orthologues are expressed in cell clusters containing epidermis marker genes (Fig. 1b and Supplementary Fig. 5b). The observed compartmentalization of IPAP-specific iridoid biosynthesis and epidermis-specific downstream biosynthesis is shown here for a *Rubiaceae* species and is identical to that found in *C. roseus* (*Apocynaceae*), suggesting that this spatial organization is conserved between these iridoid-producing families<sup>12,27–30</sup>.

With these datasets in hand, we filtered the transcript lists generated from bulk tissue RNA-seq and snRNA-seq co-expression analyses for high absolute expression levels (counts per million (CPM) > 50 in young leaves; cluster average expression > 1 within the IPAP cell cluster), based on the assumption that the cyclase gene would be highly expressed. Combining the filtered lists provided 13 gene candidates (Fig. 1c and Extended Data Fig. 1a). These 13 genes included all seven IPAP-specific iridoid biosynthesis genes, the two rate-limiting genes *DXS* and *DXR* of the 2-C-methyl-D-erythritol 4-phosphate pathway that makes the GPP precursor isopentenyl phosphate, and only four uncharacterized genes (Fig. 1c and Extended Data Fig. 1a). It is worth noting that when less stringent absolute expression value cut-offs were applied, the overlaid list also contained orthologues of the three known *C. roseus* basic helix–loop–helix iridoid biosynthesis transcriptional regulators, highlighting that this strategy can also be used to identify cell-type-specific transcriptional regulators (Extended Data Fig. 1b)<sup>31–33</sup>.

We had previously developed a mass spectrometry-based detection method for 7S-*cis-trans* nepetalactol-derived iridoids in transfected *N. benthamiana* leaves<sup>22</sup>. We capitalized on this method to screen the activity of the cyclase gene candidates by assaying them in the context of upstream (*GPPS*, *GES*, *G8H*, *8HGO*, *ISY*) and downstream (7-deoxyloganin acid glucosyl transferase (*ZDLGT*), *ZDLH*) iridoid biosynthetic pathway genes. The cyclase candidates were expressed with *C. ipecacuanha* orthologues of iridoid biosynthetic genes predicted to generate loganic acid, an iridoid derived from 7S-*cis-trans* nepetalactol, in *N. benthamiana*. Inclusion of one of the candidates<sup>34</sup>, which was functionally annotated as a MES, resulted in the efficient production of loganic acid (Fig. 1e). The enzyme was thus named Iridoid cyclase (ICYC). This protein is entirely unrelated to the *Nepeta*-specific cyclases (nepetalactol-related short chain reductases and MLPLs), clearly indicating that iridoid cyclization arose convergently at least once. We systematically compared the *C. ipecacuanha* ICYC (CiICYC) sequence against available data for 20 asterid clades<sup>35</sup> and identified ICYC orthologues in all clades reported to produce iridoids with the exception of the Apiales and Aquifoliales iridoid-producing genera *Griselinia* and *Helwingia*, respectively<sup>36,37</sup>. Analogously, an ICYC orthologue appeared to be absent in non-iridoid-producing clades, as well as in *Nepeta* (Fig. 1d; see Supplementary Fig. 6 for a tree with all orthologues). It is worth noting that *ICYC* is located next to the iridoid pathway gene *G8H*, forming a small biosynthetic gene cluster that is conserved in all iridoid-producing orders for which genome data is available (Fig. 1d).

In addition to CiICYC (lamiids, Gentianales, Rubiaceae), we selected orthologues from six additional species representing different asterid orders and families: the 7S-*cis-trans* nepetalactol-producing species *A. salviifolium* (*AsICYC*, Cornales), *Vaccinium corymbosum* blueberry (*VcICYC*, Ericales), *Lonicera japonica* (*LjICYC*, campanulids, Dipsacales) and *C. roseus* (*CrICYC*, lamiids, Gentianales, Apocynaceae), and the 7R-*cis-cis* nepetalactol-producing Lamiales species *Sesamum indicum* (*SiICYC*, Pedaliaceae) and *Antirrhinum majus* (*AmICYC*, Plantaginaceae)<sup>15,26,38–40</sup>. Inclusion of each of these genes resulted in the effective reconstitution of loganic acid biosynthesis in *N. benthamiana* (Fig. 1e and Supplementary Fig. 7). We further used these cyclases to reconstitute the complete downstream *C. ipecacuanha* and *A. salviifolium* iridoid pathways (Supplementary Fig. 8). Finally, to confirm ICYC activity in a native plant, we performed virus-induced gene silencing (VIGS) of *CrICYC* in *C. roseus* (a technique not available in *C. ipecacuanha*) and detected reduced iridoid (secologanin) content, consistent with the proposed iridoid cyclization function (Extended Data Fig. 2).

To rigorously characterize the direct product of ICYC, we assayed seven recombinantly produced ICYC orthologues together with 8-oxogeranial and the previously characterized 7S-*ISY* from *C. roseus* (*CrISY*) and 7R-*ISY* from *A. majus* (*AmISY*) (Fig. 2 and Supplementary Figs. 9 and 10)<sup>13,15</sup>. *CrISY* and *AmISY* reduce 8-oxogeranial to form the unstable intermediate S- or R-8-oxocitronellyl enol, respectively, which, in the absence of a cyclase, spontaneously forms a small amount of *cis-trans* nepetalactol, along with a wide variety of side products (Fig. 2a; see Supplementary Fig. 9 for detailed description of all spontaneous products)<sup>14,15</sup>. When *CrISY* was assayed together with any of the ICYC orthologues, 7S-*cis-trans* (7S, 4aS, 7aS) nepetalactol accumulation was substantially increased compared to when *CrISY* was assayed alone (Fig. 2b). When *AmISY* was assayed with ICYC orthologues, a new peak appeared (Fig. 2c). We confirmed the identity of this ICYC-specific product as 7R-*cis-cis* (7S, 4aS, 7aS) nepetalactol by chemical oxidation and comparison with an authentic standard of 7R-*cis-cis* nepetalactone (Fig. 2d–f)<sup>23</sup>. In addition, we observed comparable results when we assayed the ICYC orthologues directly with S-8-oxocitronellal and R-8-oxocitronellal at high concentrations of general acid (0.5 M 3-(*N*-morpholino)propanesulfonic acid (MOPS)), conditions that promote partial tautomerization to the putative cyclase substrate, 8-oxocitronellyl enol<sup>14</sup> (Supplementary Fig. 11).

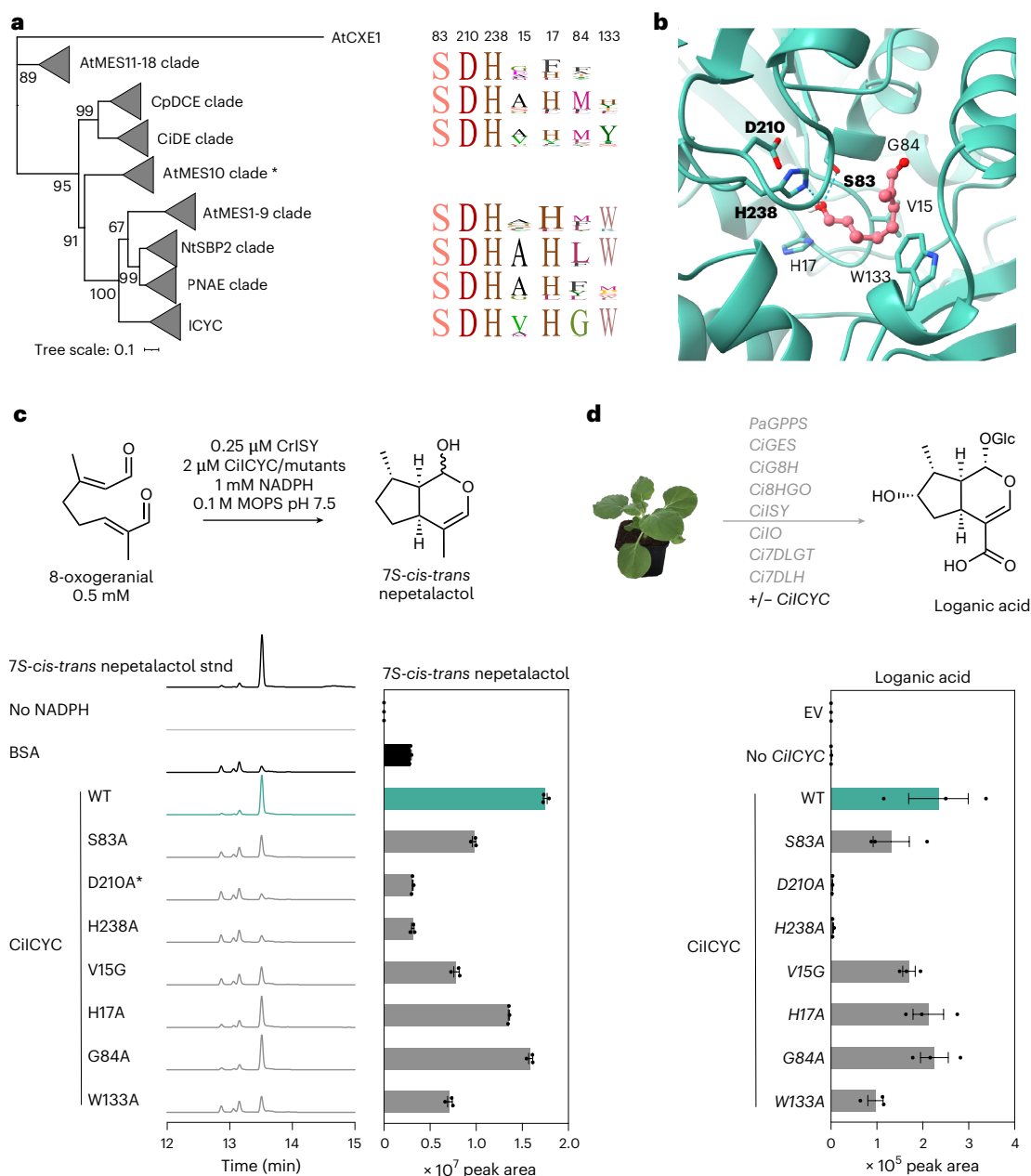


**Fig. 2 | In vitro activity assays reveal stereoselectivity of ICYCs.** **a**, Scheme depicting reaction products of 7*S* and 7*R* stereo-selective ISY together with ICYC. **b,c**, Assays were performed as indicated and analysed by GC-MS. BSA was used as a negative control. Displayed chromatograms are total ion chromatograms of one representative replicate. See Supplementary Fig. 9 for detailed descriptions of chromatograms including side products and Supplementary Fig. 10 for fragmentation patterns of standards and enzymatic products. Bar graphs depict peak areas from  $N = 3$  replicates of nepetalactol (shaded in grey); error bars are standard error of the mean (s.e.m.). **b**, Assays of ICYC orthologues with the 7*S* selective *C. roseus* ISY (CrISY) show a dramatic increase in 7*S*-*cis-trans* nepetalactol in the presence of ICYC orthologues. **c**, Assays of ICYC orthologues with the 7*R* selective *A. majus* ISY (AmISY) reveal appearance of a peak consistent

with 7*R*-*cis-cis* nepetalactol in the presence of ICYC. The asterisk indicates that these side products are the 7*R* enantiomers that coelute with the 7*S* series when using an achiral stationary phase as used here (Supplementary Fig. 9). **d**, To confirm the identity of the AmISY–AmICYC enzymatic product, five enzymatic reactions were pooled and chemically oxidized to 7*R*-*cis-cis* nepetalactone (Methods). **e**, Chiral GC-MS analysis of the chemically oxidized product showed that it had the same retention time and the identical mass (extracted ion chromatogram for nepetalactone mass  $m/z$  166.1) as the authentic standard<sup>23</sup> on a chiral GC column. **f**, The MS fragmentation pattern of the oxidized enzymatic product was identical to the standard confirming that the enzymatic product is indeed 7*R*-*cis-cis* nepetalactol. Ctrl, control.

Taken together, these results show that ICYCs exclusively produce 4*aS*, 7*aR* nepetalactol regardless of the stereochemistry of the C-7-position. Strikingly, while all ICYC orthologues generate 7*S*-*cis-trans* nepetalactol from S-8-oxocitronellal, 7*R*-*cis-cis* nepetalactol is most efficiently

generated when ICYC orthologues from the Lamiales species known to produce this stereoisomer (that is, AmICYC and SilICYC) are used. This observation suggests that enzymes in these species might have specialized to adapt to the 7*R* substrate. We also assayed the



**Fig. 3 | ICYC is a member of the MES family.** **a**, A condensed phylogenetic tree showing that ICYC is related to MESs that esterify substrates in plant hormone activation pathways and in alkaloid biosynthesis. Each clade is named after a member found within the respective clade. See Supplementary Fig. 12 for an extended version of this tree. Sequence data highlight the conservation of the catalytic triad and other amino acid residues that were subjected to mutation in CiICYC. The asterisk indicates that AtMES10 was a single-member clade, and thus no sequence logos were generated. At, *Arabidopsis thaliana*; CXE, carboxylesterases; CiDE, *C. ipeacuanha* deacetyl(iso)ipecoside esterase; CpDCE, *Cinchona pubescens* dihydrocorynantheine aldehyde esterase; RSPNAE, *Rauwolfia serpentina* polyneuridine aldehyde esterase. **b**, Putative active site of an alphaFold3 structural model of CiICYC with the docked (AutoDock Vina) substrate 5-8-oxocitronellyl enol substrate. Indicated amino acids were chosen

for mutagenesis studies because they are part of the catalytic triad (in bold) or are in proximity to the docked substrate and are conserved in ICYC orthologues. **c**, In vitro assays show that D210A and H238A ICYC mutants yield a reaction profile that is identical to the negative control. S83A, V15G and W133A mutants show reduced levels of nepetalactol compared to wild-type ICYC, whereas H17A and G84A mutants behaved similarly to wild-type (WT) ICYC. The asterisk indicates that the D210A mutant exhibited poor solubility, and thus the assays with this mutant should be interpreted cautiously. Bar graphs depict normalized peak areas from  $N = 3$  replicates of nepetalactol (shaded in grey); error bars are s.e.m. **d**, Reconstitution of loganic acid biosynthesis in *N. benthamiana* with CiICYC and CiICYC mutants yielded results that are consistent with results from in vitro assays shown in **c**. Bar graphs depict normalized peak areas from  $N = 3$  biological replicates; error bars are s.e.m. EV, empty vector.

product profile when ICYC was added after ISY reduced 8-oxogeranial to 8-oxocitronellyl enol (Extended Data Fig. 3a). Under these conditions, no increase in nepetalactol is observed, suggesting that 8-oxocitronellyl enol must be immediately transferred to ICYC after it is formed by ISY. In support of this hypothesis, we found that ISY and ICYC interact (as measured by split luciferase assays) (Extended Data Fig. 3b).

However, this observation does not definitively show that substrate channelling, which would protect the reactive enol species, in fact takes place.

ICYC is a methyl esterase (MES)-type  $\alpha/\beta$  hydrolase. These MESs are ubiquitously found in plants but have not previously been associated with non-esterase functions such as cyclization. Many MES

family members play roles in activating the plant hormones methyl jasmonate, methyl salicylate or indole-3-acetic acid methyl ester via demethylation<sup>41–43</sup>. Other members act as esterases in monoterpene indole or ipecac alkaloid biosynthetic pathways<sup>26,44–48</sup>. All  $\alpha/\beta$  hydrolases use a conserved catalytic triad (serine, aspartate/glutamate and histidine) along with a glycine-rich oxyanion hole to catalyse substrate hydrolysis (Fig. 3a,b)<sup>49,50</sup>. Although ICYC orthologues form a phylogenetic clade well separated from all other MESs, indicating a monophyletic origin (Fig. 3a and Supplementary Fig. 12), the catalytic triad and oxyanion hole motifs are still present, as in canonical esterases (Fig. 3a and Supplementary Fig. 13). No obvious differences were noted within the active sites of ICYCs from species that produce 7R nepetalactol compared to those that produce 7S nepetalactol (Supplementary Fig. 14). All ICYC orthologues showed esterase activity when tested with the model substrate 4-nitrophenyl acetate, indicating that these catalytic motifs remain functional (Extended Data Fig. 4). To assess which residues are responsible for catalysis of cyclization, we docked the S-8-oxocitronellyl enol substrate to a CiICYC alphaFold3 model with AutoDock Vina to pinpoint residues in the binding pocket. We also used our phylogenetic analysis to select amino acid residues that are conserved in ICYC orthologues (Fig. 3a,b). Unsurprisingly, mutations of any of the catalytic triad amino acid in CiICYC (S83A, D210A, H238A) abolished esterase activity (Extended Data Fig. 4c)<sup>51</sup>. D210A was largely insoluble and could therefore not be assayed reliably. When these mutants were tested for cyclase activity in vitro (Fig. 3c) and in *N. benthamiana* (Fig. 3d), D210A and H238A mutants were inactive, whereas the S83A mutant exhibited some cyclization activity. Out of the four additional mutants within the putative cyclase active site, V15G and W133A led to reduced amounts of the nepetalactol product (Fig. 3c,d). Although more extensive experimentation is required to understand the mechanism of this unusual cyclization (Supplementary Fig. 15), we speculate that these residues shape the binding pocket of ICYC, allowing 8-oxocitronellyl enol to adopt the conformation required for cyclization. The amino acid H238, the mutation of which abolished detectable activity, may play a role in orienting the substrate. Alternatively, this residue could interact with the enol moiety of 8-oxocitronellyl enol, which could in turn activate the substrate to undergo stepwise cyclization (Supplementary Fig. 15d)<sup>52</sup>.

The discovery of these stereoselective ICYC orthologues fully explains the dominance of 7S *cis-trans* and 7R *cis-cis* nepetalactol derived iridoids in the asterids. This work clearly shows that ICYC was recruited from widely present MESs to perform this unusual cyclization reaction and has since been evolutionarily conserved throughout iridoid-producing asterids. It is worth noting that an apparent evolutionary loss of ICYC in the *Nepetoideae* may have facilitated the convergent evolution of cyclases with different stereoselectivities in *Nepeta*. The discovery of ICYC is an excellent starting point to mechanistically understand this unusual cyclization reaction<sup>52,53</sup>. Finally, this discovery also unlocks the possibility for formation of previously inaccessible iridoid stereoisomers, which will enable metabolic engineering for sustainable production of valuable iridoid and iridoid-derived compounds.

## Methods

### Plant sampling and RNA extraction

*C. ipecacuanha* and *A. salviifolium* plants were previously obtained<sup>26</sup>. Plants labelled with different numbers refer to independent individual plants. RNA-seq data from plants labelled as ‘plant 1’ was previously published (National Center for Biotechnology Information (NCBI) BioProject PRJNA1169657). All plants were grown under the following conditions: 12/12 h light/dark, 28–30 °C/24–26 °C temperature, 70–80% humidity. Plants for Illumina sequencing and expression analyses (see Supplementary Fig. 2a,b for photographs) were the following: *C. ipecacuanha* plants labelled as ‘plant 2’ and ‘plant 3’ were both 1.5 years old; *A. salviifolium* plants labelled as ‘plant 2’ and ‘plant 3’

were 2.5 and 4 years old, respectively. Dissected tissues from *C. ipecacuanha* plants 2 and 3 and *A. salviifolium* plant 2 were flash frozen and shipped on dry ice to FutureGenomics for RNA extraction and sequencing (see below). Additional tissues were collected for Oxford Nanopore Technologies (ONT) sequencing for genome annotation (young leaves, mature leaves, green stem and roots of 2-year-old *C. ipecacuanha* plants; leaf buds, young leaves, mature leaves and roots of 3-year-old *A. salviifolium* plants). RNA of samples for ONT sequencing and *A. salviifolium* plant 3 samples was extracted in-house as follows. Dissected tissues were immediately flash frozen into liquid nitrogen and ground in liquid nitrogen using an IKA A11 basic analytical mill or mortar and pestle. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer’s instructions, including on column DNase digest. RNA concentrations and purity were determined with a Nanophotometer N60 (Implen).

### RNA Illumina sequencing

Illumina sequencing of *C. ipecacuanha* plants 2 and 3 and *A. salviifolium* plant 2 was performed at FutureGenomics as follows. Flash-frozen plant tissues were powdered in liquid nitrogen using mortar and pestle, and RNA was extracted using Zymo Quick-RNA Plant Kit (Zymo Research). The quality of the RNA was analysed using RNA ScreenTape on an Agilent 4200 TapeStation System (Agilent Technologies Netherlands BV), and the quantity was measured using a Qubit 3.0 Fluorometer (Life Technologies Europe BV). Illumina RNAseq libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit Illumina and NEBNext Multiplex Oligos for Illumina (New England Biolabs). The quality of the libraries was checked using Agilent D1000 screentape on an Agilent 4200 TapeStation System. RNA library paired-end sequencing (2 × 150 bp) was performed using Illumina’s NovaSeq 6000 technology. Two rounds of sequencing were carried out to reach the desired output target.

Illumina sequencing of *A. salviifolium* plant 3 was performed at Novogene. RNA samples were shipped to Novogene where messenger RNA library preparation and sequencing were performed according to the company’s standard protocol for mRNA sequencing. RNA integrity and quantitation were assessed on a Bioanalyzer (Agilent Technologies). All samples were above the required minimum RNA integrity number value. Sequencing was performed on an Illumina NovaSeq X Plus PE150 platform with a data output target of 9 G of raw data.

### ONT full-length complementary DNA sequencing

mRNA was purified from the total RNA using the Dynabeads mRNA Purification Kit (Invitrogen) and input into the ONT SQK-PCS111 kit to generate full length complementary DNA libraries. Resultant libraries were sequenced on a MIN106 Rev. D flowcell before being basecalled using Guppy (v6.5.7) (<https://nanoporetech.com/software/other/guppy>) using the super high accuracy model (dna\_r9.4.1\_450bps\_sup.cfg) and the parameters -q 0, -trim\_strategy none and --calib-detect.

### Genome sequencing and assembly

Nuclei were extracted from young leaves using a nuclei isolation protocol<sup>54</sup>. The nuclear pellet was resuspended in Qiagen buffer G2 with RNaseA and proteinaseK, and DNA extraction was further continued according to the instructions of the Qiagen Genomic Tip/100 G protocol (Qiagen). The quality of the DNA was analysed using Genomic DNA ScreenTape on an Agilent 4200 TapeStation System (Agilent Technologies Netherlands BV), and the quantity was measured using a Qubit 3.0 Fluorometer (Life Technologies Europe BV). Nanopore sequencing libraries were prepared using the Ligation Sequencing Kit V14 (SQK-LSK114) according to the manufacturer’s instructions (ONT). Each library was run on an R10.4.1 PromethION flowcell (FLO-PRO114M; ONT) and reloaded on a daily basis after a nuclease flush with Flow Cell wash kit (EXP-WSH004). Super-accuracy base calling was done using Guppy 6.3.2. Illumina DNaseq libraries were prepared using the

Nextera Flex kit according to the manufacturer's instructions (Illumina) and were sequenced in paired-end mode (2 × 150 bp) using Illumina's NovaSeq 6000 technology.

ONT genomic reads were assembled using Flye (v.2.9.1), with settings 'overlap 10K, error rate 0.025, no-alt-contigs'<sup>55</sup>. The genome contigs were polished using nanopore reads by Medaka (<https://github.com/nanoporetech/medaka>) and then polished twice using Illumina reads by Pilon (v.1.23)<sup>56</sup>. The polished genome sequence was then collapsed using Purge dups v1.2.6<sup>57</sup>.

### Genome annotation

The genome assemblies were repeat masked by first creating a custom repeat library (CRL) for each genome. Repeats were first identified with RepeatModeler<sup>58</sup> (v2.03), and protein-coding genes were filtered out from the repeat database using ProtExcluder<sup>59</sup> (v1.2) to create a CRL. The CRL was then combined with Viridiplantae repeats from RepBase (v20150807) to generate the final CRL for each genome. Each genome assembly was repeat-masked using the respective final CRL and RepeatMasker<sup>60</sup> (v4.1.2-p1) using the parameters -e ncbi -s -nolow -no\_is -gff.

RNA-seq libraries were processed for genome annotation by first cleaning with Cutadapt<sup>61</sup> (v2.10) using a minimum length of 100 nt and quality cut-off of 10 then aligning the cleaned reads to the respective genome using HISAT2<sup>62</sup> (2.1.0). ONT cDNA reads were processed with Pychopper (<https://github.com/epi2me-labs/pychopper>) (v2.7.10), and trimmed reads greater than 500 nt were aligned to the respective genome using minimap2<sup>63</sup> (v2.17-r941) with a maximum intron length of 5,000 nt. The aligned RNA-seq and ONT cDNA reads were each assembled using Stringtie<sup>64</sup> (v2.2.1), and transcripts less than 500 nt were removed.

The initial gene models for each genome were created using BRAKER2<sup>65</sup> (v2.1.6) using the soft-masked genome assemblies and the aligned RNA-seq libraries as hints. The gene models were then refined using two rounds of PASA2<sup>66</sup> (v2.5.2) to create a working gene model set for each genome. High-confidence gene models were identified from each working gene model by filtering out gene models without expression evidence or a PFAM domain match, or were a partial gene model or contained an interior stop codon. Functional annotation was assigned by searching the working gene models proteins against The *Arabidopsis* Information Resource (TAIR)<sup>67</sup> (v10) database and the Swiss-Prot plant proteins (release 2015\_08) database using BLASTP<sup>68</sup> (v2.12.0) and the PFAM<sup>69</sup> (v35.0) database using PfamScan<sup>70</sup> (v1.6) and assigning the annotation based on the first significant hit.

### Single-nuclei RNA-seq of *C. ipeacacuanha* young leaves

Nuclei were isolated following the protocol outlined by ref. 71. For *C. ipeacacuanha* young leaves 0.01% Triton X-100 was used, and RNase inhibitor (Sigma Protector RNase Inhibitor, catalogue number 3335402001; SigmaAldrich) was added to the nuclei isolation buffer for a final concentration of 0.5 U μl<sup>-1</sup>. Nuclei were stained with DAPI (4',6'-diamidino-2-phenylindole) and sorted using fluorescence activated cell sorting. Nuclei were concentrated by spinning at 300 g for 5 min. Concentrated nuclei were used for single-cell RNA-seq library construction using the PIPseq T20 v4.0Plus Kit (Illumina) with 1 μl of additional RNase inhibitor.

### Single-cell transcriptomics and co-expression analyses

Single-nuclei RNA-seq reads were processed using the 'barcode' command from pipseeker-v3.1.3 (Illumina). The pipseeker processed fastq files, and the generated barcode whitelist was used as input into the STARsolo (v2.7.10b)<sup>72</sup> alignment program. The following parameters were used: --alignIntronMax 5000, --soloUMIlen 12, --soloCellFilter EmptyDrops\_CR, --soloFeatures GeneFull, --soloMultiMappers EM and --soloType CB\_UMI\_Simple. Seurat v4.3.0.1 was used for downstream analysis. Samples were filtered to retain high-quality cells by removing cells with less than 300 genes or more than 10,000 genes and less than

500 UMIs (unique molecular identifiers) or more than 30,000 UMIs. Samples were also run through DoubletFinder v2.0.4<sup>73</sup> to remove suspected doublets. Reciprocal principal component analysis was used to integrate the two replicates together using the top 3,000 variable genes. The top 60 principal components were used with a resolution parameter of 0.5 to calculate the uniform manifold approximation and projection.

SimpleTidy\_GeneCoEx v2.0.0<sup>74</sup> was used for co-expression analysis. An *R* value cut-off of 0.7 with a resolution parameter of 3 was used to generate 27 modules containing 5 or more genes, which comprised 29,765 genes total. Module 16 contained IPAP-specific genes of interest (1,496 genes in the module).

### Mapping of bulk RNA-seq to *A. salviifolium* and *C. ipeacacuanha* assembled genomes

Adapter-cleaved raw fastq files were received, and reads were quality checked with FastQC (Galaxy version 0.73) and trimmed using Trimmomatic (Galaxy version 0.38.1)<sup>75</sup> on an in-house Galaxy server<sup>76</sup>. Reads belonging to the same sample but obtained in two sequencing rounds were concatenated. Reads were mapped to genomes using CLC Genomics workbench 21.0.4 (Qiagen) with these parameters: mismatch cost, 2; insertion cost, 3; deletion cost, 3; length fraction, 0.85; similarity fraction, 0.9; auto-detect paired distances, on; and maximum number of hits for a read, 20. Expression values are unique counts per gene. trimmed mean of M values normalized CPM values were used for downstream analyses. To identify iridoid pathway genes in *C. ipeacacuanha* and *A. salviifolium*, amino acid sequences of the characterized iridoid pathway enzymes from *C. roseus* were blasted (tblastn) against the transcript working models derived from the genomes. The highest blast hit for each gene was chosen for expression analysis and cloning (Supplementary Fig. 3a). Blasting of enzymes known to perform cyclizations in *Nepeta*<sup>19,24</sup> yielded comparably poorly conserved sequences (maximum 59% identity) that were not co-expressed with iridoid pathway genes and were thus not considered as orthologues. *C. ipeacacuanha* tissue-specific co-expression analysis was performed on an expression atlas containing data from all tissues from plants 1–3. *CiSY* expression was used as a bait for Pearson correlation.

### Gene cloning

Complementary DNA was prepared from total RNA of *A. salviifolium* leaf buds and roots, *C. ipeacacuanha* young leaves and *C. roseus* leaves (extracted as described above) using the iscript cDNA Synthesis Kit (Biorad) according to manufacturer's instructions. *NmMLPL* and *PaGPPS* genes had been cloned previously<sup>22,24</sup>. Genes from all other species and *CiCYC* protein mutants were obtained as synthetic sequences from Twist Biosciences. *CDS* sequences were amplified with the Q5 High-Fidelity 2X Master Mix (New England Biolabs) using cDNA or synthetic fragments as templates and gene-specific primers containing overhangs for In-Phusion cloning (Supplementary Table 8). Amplified sequences were gel-purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) and cloned using the 5x In-Fusion Snap Assembly Master Mix (TaKaRa Bio). For pathway reconstitution in *N. benthamiana*, coding regions were inserted into a modified 3Q1 vector (contains *UBQ10* promoter and terminator from *Solanum lycopersicum*<sup>77</sup>) previously digested with BsaI-HF v2 (New England Biolabs, NEB). For expression in *Escherichia coli*, sequences were cloned into the pOPINF vector<sup>78</sup> previously digested with KpnI-HF and HindIII-HF (NEB). ICYCs are predicted to be cytosolic and do not contain any transit peptides to be cleaved off for recombinant protein production (predicted by DeepLoc 2.1 (ref. 79)). For split-luciferase assays, *CiCYC* and *CiSY* were cloned into KpnI-HF and Sall-HF (NEB) digested pCAMBIA1300-NLuc (N-terminal luciferase) and KpnI-HF and PstI-HF (NEB) digested pCAMBIA1300-CLuc (C-terminal luciferase)<sup>80</sup>. Cloning reactions were transformed into heat shock competent *E. coli* TOP10 and grown overnight in a 37 °C incubator on Luria-Bertani

(LB) agar plates containing the respective antibiotics. Plasmids were isolated from overnight cultures of single colonies using the Wizard Plus SV Minipreps DNA Purification System kit (Promega), and inserted sequences were confirmed by Sanger sequencing.

### ***Agrobacterium tumefaciens* mediated transient expression in *N. benthamiana***

*A. tumefaciens* GV3101 cells were transformed through electroporation, recovered in YEB without antibiotics and incubated on YEB plates containing antibiotics (rifampicin and gentamycin and the appropriate antibiotic for plasmid selection) at 28 °C for 48 h. Colony PCR was done on single colonies to confirm the presence of plasmids. Positive colonies were grown in liquid YEB for 24 h. From these cultures glycerol stocks were prepared and stored at -80 °C. Three- to four-week-old *N. benthamiana* plants (grown in a greenhouse with 16 h/8 h light/dark, 23–26 °C/16–22 °C temperature and 40–70% humidity) were used for agroinfiltration<sup>26,81</sup>: cells from glycerol stocks were spread on YEB plates containing antibiotics and 100 µM acetosyringone and grown for 24 h until a visible layer of bacteria appeared. The bacteria were transferred to 1–2 ml of infiltration medium (10 mM MES, 10 mM MgCl<sub>2</sub>, 100 µM acetosyringone, pH 5.7) and gently resuspended, and the optical density at 600 nm (OD<sub>600</sub>) was measured in 1:10 dilutions using an Implen OD600 DiluPhotometer. For pathway reconstitution experiments, strains were mixed and diluted in infiltration buffer to OD<sub>600</sub> = 0.1 per strain. For split-luciferase assays (see below), the strains were mixed at OD<sub>600</sub> = 0.13 per strain. Strains harbouring constructs with NLuc or CLuc fused to CiiCYC or CiiSY or empty vectors containing free NLuc or CLuc as controls were infiltrated in combinations as indicated. A strain harbouring a construct with the *p19* gene was co-infiltrated in all cases. The culture mixtures were infiltrated into *N. benthamiana* leaves and grown for 5 days under grow lights (16 h/8 h light/dark). Replicates are from individual plants. For metabolite analysis, leaf material was collected 5 days after agroinfiltration by flash freezing in tubes containing metal beads.

### **Split-luciferase assays in *N. benthamiana***

Leaf disks (-1 cm) were punched from four biological replicates 3 days after infiltration and placed into a custom-made high-density polyethylene multi-well plate. The abaxial side of the leaf was facing up. Then 200 µl of 0.5 mM D-luciferin (Promega) was added to the leaf disks in each well. The plate was imaged with a NightSHADE LB 985 (Berthold Technologies) with luminescence emission at 0.1 s (wavelength filter 650 nm, 10% intensity) and 8 × 8 pixel binning. Images were taken after 15 and 20 min of incubation with luciferin in the dark with 10 s and 2 s exposure times, respectively. The pictures were analysed with indigo-OTM 1.4 software (Berthold Technologies) and a scale from 5,000 to 65,000 counts per second was applied.

### **VIGS in *C. roseus***

VIGS was performed according to an established method<sup>29</sup>. Briefly, 300 bp target regions of the coding regions of *C. roseus* *ICYC* and *ISY* genes (transcripts CRO\_01G006740.1 and CRO\_07G007680.1 at <https://doi.org/10.5061/dryad.d2547d851>, therein *cro\_v3.gene\_models.cds.fa* ([https://datadryad.org/downloads/file\\_stream/2121644](https://datadryad.org/downloads/file_stream/2121644))) were selected using the SGN VIGS tool to avoid off-target gene silencing (<https://vigs.solgenomics.net/>)<sup>82</sup>. Genomic DNA was extracted from *C. roseus* leaves with the DNeasy Plant Mini Kit (Qiagen). Target region fragments were PCR-amplified from gDNA with in-fusion cloning overhangs using Phusion High-Fidelity DNA Polymerase (ThermoFisher) and specific primers (Supplementary Table 7). The obtained PCR fragments were cloned in the BamHI and XhoI (NEB) digested VIGS vector pTRV2-MgChl<sup>83</sup> using the In-Fusion Snap Assembly Master Mix (TaKaRa Bio).

*A. tumefaciens* GV3101 was transformed by electroporation as described above. For inoculation, *A. tumefaciens* GV3101 with plasmid pTRV1<sup>84</sup> and *A. tumefaciens* GV3101 carrying pTRV2-MgChl,

pTRV2-CrICYC or pTRV2-CrISY were grown overnight in a rotary shaker at 28 °C and 300 rpm in each 10 ml of LB medium supplemented with 50 mg l<sup>-1</sup> kanamycin, 25 mg l<sup>-1</sup> gentamicin and 100 mg l<sup>-1</sup> rifampicin to an OD<sub>600</sub> of ~2. Cultures were centrifuged for 10 min at 1,800 g, and pellets resuspended in infiltration buffer (100 µM acetosyringone, 10 mM NaCl and 1.75 mM CaCl<sub>2</sub>) to an OD<sub>600</sub> of 2. After 2 h of incubation on a rotary shaker at room temperature and 60 rpm, 450 µl of each bacterial strain containing a pTRV2 plasmid were mixed with the same volume of the bacterial strain with pTRV1. VIGS inoculation was performed by pipetting 10 µl of the mixed bacterial suspension between plant stem and petiole of the first true leaf of a 30-day-old *C. roseus* cultivar 'Atlantis Burgundy Halo' plant (grown in a growth chamber at 16/8 h light/dark 23 °C/21 °C temperature, 50% humidity). The stem was pierced with a Ø 0.40 × 25 mm Sterican needle ([www.bbBraun.com](http://www.bbBraun.com)) twice through the bacterial suspension drop. Six plants per pTRV2 construct were inoculated. Plants inoculated with the pTRV2-MgChl strain served as controls. After VIGS inoculation, plants were grown under the same conditions as described above. Yellow tissues (due to co-silencing of *MAGNESIUM CHELATASE SUBUNIT H* gene) were collected 3 weeks after inoculation.

### **Quantitative PCR**

Total RNA of *C. roseus* silenced tissues was extracted using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions, including on column DNase digest. RNA concentrations were measured with a Nanophotometer N60 (Implen). mRNA was reverse-transcribed with 0.5 µg total RNA as input using the iscript cDNA synthesis kit (Bio-rad) according to the manufacturer's instructions. Complementary DNA was then diluted 1:8 in water, and 2 µl of the dilution was used in each quantitative PCR (qPCR) reaction of 10 µl total volume in a 96-well plate with 333 nM of each primer (Supplementary Table 8) and 5 µl of PowerUp SYBR Green Master mix (Applied Biosystems) according to manufacturer's instructions. Amplification was done on a QuantStudio 1 (Applied Biosystems) qPCR machine with the following cycling conditions: 50 °C for 2 min, 95 °C for 2 min, followed by 40 cycles of 95 °C for 1 s and 60 °C for 30 s (ramp rate was 1.6 °C s<sup>-1</sup> in all cases). Amplification was followed by a melting curve (0.15 °C s<sup>-1</sup> from 60 °C to 95 °C) to confirm the presence of single amplification products. Data were analysed using the  $\Delta\Delta C_t$  method and normalized to the expression of the established reference gene *N2227*<sup>85</sup>. Graphs and statistics were done in Prism Graphpad 10.4.1.

### **Metabolite extraction**

Leaf material of agroinfiltrated *N. benthamiana* or silenced *C. roseus* was ground using two 4 mm metal beads and a TissueLyser (Qiagen) with pre-cooled adapters and extracted with 30 µl per mg of fresh material of 70% MeOH containing 0.1% formic acid and 1 µM harpagoside as internal standard. Samples were sonicated for 10 min, incubated on a rotator for 15 min and centrifuged at 18,000 g for 15 min. The supernatants were filtered through a 0.45 µm low binding hydrophilic PTFE filter plate (MultiScreen Solvintert 96, Merck-Millipore) into a 96-well Microtiter Plate (SureSTART WebSeal, Thermo Scientific) according to manufacturer's instructions. Plates were sealed with Rapid Slit Seal (BioChromato) and immediately analysed with ultra-performance liquid chromatography–tandem mass spectrometry (UPLC-MS/MS).

### **Detection of iridoid glucosides by UPLC-MS/MS**

The system consisted of an UltiMate 3000 Ultra-High Performance Liquid Chromatography system (Thermo Fisher Scientific) coupled to an Impact II high-resolution Quadrupole Time-Of-Flight mass spectrometer (Bruker Daltonics). A Kinetex XB C18 (2.1 × 100 mm, 2.6 µm; 100 Å) column (Phenomenex) was set at 40 °C and 0.6 ml min<sup>-1</sup> flow rate, and 2 µl of sample was injected. The mobile phase was A:B where A was water with 0.1% formic acid and B was acetonitrile. The gradient was as follows: 5% B at 0.5 min to 30% B at 6 min. Then the column

was flushed at 100% B until 7.6 min and re-equilibrated to 5% B until 10 min. Ionization was performed in negative electrospray ionization mode (ESI<sup>-</sup>) with 3,500 V capillary voltage and 500 V end plate offset, a nebulizer pressure of 2.5 bar with nitrogen at 250 °C and a flow of 11 l min<sup>-1</sup> as the drying gas. Acquisition was done at 12 Hz following a mass range from 100 to 1,000 *m/z* with data-dependent MS/MS, an active exclusion window of 0.2 min and a reconsideration threshold of 1.8-fold change. Fragmentation was triggered on an absolute threshold of 400 and limited to a total cycle time range of 0.5 s. For collision energy, the stepping option model (from 20 to 50 eV) was used. Recalibration of the *m/z* values took place at the start of each run using the expected cluster ion *m/z* values of a direct source infusion of sodium formate-isopropanol solution. At the first minute of each run, the liquid chromatography (LC) input was redirected to waste. During this time, the *m/z* values of the instrument were calibrated using the cluster ion *m/z* values of a sodium formate-isopropanol solution injected by direct source infusion with a 5 ml syringe connected to an external pump at a flow rate of 0.18 ml h<sup>-1</sup>. LC-MS data were collected with Bruker Compass qtofControl 5.2.109/Hystar 5.1.5.1 or Compass qtofControl 6.3/Hystar 6.0.30.0 software.

#### UPLC-MS/MS data analysis

Raw data were converted to mzml format using MSConvert 3.0.25036-69e37b6 and imported to MZmine 4.5.37 (ref. 86,87). Extracted ion chromatogram traces and MS/MS data of compounds of interest were exported from MZmine. Peak areas were calculated using the MZmine Processing Wizard and exported. Peak areas of compounds of interest were normalized to the internal standard harpagoside and converted to intensity per second. Further data analysis and construction of graphs was done in GraphPad Prism 10.4.1 for Mac OS X.

#### Commercially available chemicals and standards

Secologanin (50741), loganic acid (PHL80492), 4-nitrophenyl acetate (N8130) and D-camphor (50843) were obtained from Sigma. NADPH tetrasodium salt (10621692001) was obtained from Roche; 8-oxogeranial (D476180) and 7*S*-*cis*-*trans* nepetalactol (N390065) were purchased from Toronto Research Chemicals. Harpagoside (7471.2) was purchased from Roth.

#### Synthesized standards and substrates

Secologanic acid standard was produced through alkaline hydrolysis of secologanin<sup>26,88</sup>. Secologanin was incubated with 0.1 M NaOH (40 µl per 1 mg secologanin) for 5 h and then neutralized with HCl. The completeness of the reaction was confirmed through LC-MS analysis and the solution stored at -25 °C. *S*-8-Oxocitronellal, *R*-8-oxocitronellal and 7*R*-*cis*-*cis* nepetalactone were previously obtained<sup>13,14,23</sup>.

#### Recombinant protein production and purification

Expression and purification were performed according to an established method with modifications<sup>89</sup>. *E. coli* SoluBL21 (DE3) was transformed by heat shock with pOPINF constructs. Pre-cultures were inoculated from single colonies, grown overnight at 37 °C and used to inoculate 100 ml 2× YT medium (500 ml in the case of CiiCYC\_D210A mutant). Cultures were grown at 37 °C until OD<sub>600</sub> 0.5–0.6 was reached, cooled to room temperature and induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside. After induction, the cultures were shifted to 18 °C over-night and collected the next day by centrifugation. Cell pellets were lysed on ice for 30 min with approximately 6 ml lysis buffer per 1 g of cell pellet (50 mM Tris–HCl pH8, 50 mM glycine, 5% glycerol, 500 mM NaCl, 20 mM imidazole, 0.2 mg ml<sup>-1</sup> lysozyme, 1 tablet per 50 ml of complete EDTA free protease inhibitor (Roche)) and sonicated for 2.5 min (2 s on, 3 s off) on ice (Bandelin UW 2070). Supernatants were incubated with gentle shaking in falcon tubes with 250 µl Ni-NTA Agarose (Qiagen) for 1 h at 4 °C to allow binding of His-tagged proteins. Slurry was pelleted gently by centrifugation at 1,000 g for 30 s. The

supernatant was removed, and the slurry was washed three times with ice-cold wash buffer (50 mM Tris–HCl pH 8, 50 mM glycine, 5% glycerol, 500 mM NaCl, 20 mM imidazole) by inversion, centrifugation and removal of supernatant. Proteins were eluted using elution buffer (as wash buffer but containing 500 mM imidazole). Elution fractions were concentrated, and buffer was changed to storage buffer (20 mM HEPES, 150 mM NaCl, pH 7.5) using Amicon Ultra Centrifugal Filters (Millipore) with 10 kDa molecular weight cut-off according to manufacturer's instructions. Purity was assessed through SDS–PAGE, and concentration was determined using the extinction coefficient and by measuring the absorbance at 280 nm. Proteins were flash frozen in small aliquots in liquid nitrogen and stored at -70 °C.

#### In vitro enzyme assays

With 8-oxogeranial as a substrate, enzymes were assayed under the following conditions: 100 mM MOPS pH 7.5, 0.5 mM 8-oxogeranial, 1 mM NADPH, 0.25 µM CrISY or 0.5 µM AmISY, and 2 µM ICYC, NmMLPL or bovine serum albumin (BSA) as control as indicated. The reactions were set up in a 100 µl total volume and started by the addition of NADPH. Reactions were incubated for 3 h at 30 °C and 400 rpm. For sequential reactions (Extended Data Fig. 3), ICYC, NmMLPL or BSA were added after 1.5 h of incubation with CrISY, and incubation was continued for an additional 1.5 h. For assays with *S*- or *R*-8-oxocitronellal, conditions were as follows: 500 mM MOPS pH 7.5, -0.5 mM *S*-8-oxocitronellal or -0.1 mM *R*-8-oxocitronellal, respectively, 5 µM of ICYC, NmMLPL or BSA. Reactions were incubated for 16 h at 30 °C and 400 rpm. After incubation, camphor (5 or 10 µl of a 1 mM solution in acetonitrile) was added as internal standard to each reaction and briefly vortexed. Reactions were then extracted with 200 µl ethyl acetate and vortexed thoroughly for 2 min. Layers were separated by centrifuging at 18,000g for 5 min, and 100 µl of the upper ethyl acetate layer was transferred to a glass vial. Samples were immediately analysed by gas chromatography–mass spectrometry (GC-MS).

Esterase activities were measured in a spectrophotometric assay with 4-nitrophenyl acetate<sup>90</sup>. Reactions were performed in 96-well plates (CytosOne) at 250 µl total volume. Reactions contained 100 mM HEPES pH 7, 1 mM CaCl<sub>2</sub>, 2.6 mM NaCl, 2 mM 4-nitrophenyl acetate and 0.5 µM of enzyme or BSA as control. The reactions were started by the addition of 4-nitrophenyl acetate through a multi-channel pipette, and the plate was immediately placed in a CLARIOstar Plus microplate reader (BMG Labtech). Absorbance at 405 nm was recorded every minute for a total duration of 60 min at 25 °C. Data were exported from CLARIOstar Data analysis software (MARS 3.4.1) and further analysed in Prism GraphPad 10.4.1.

#### Detection of iridoids by GC-MS

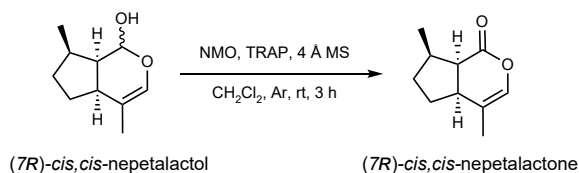
Samples were analysed on an achiral column using an Agilent 8890GC system, an Agilent 5977B GC/MSD (mass selective detector) detector and a CTC Analytics PAL RSI 120 autosampler system. A Zebtron ZB-5 Plus column (Phenomenex; internal diameter (ID) = 0.25 mm; length *L* = 30 m; film thickness = 0.25 µm) was used for separation. Samples (1 µl) were injected at 230 °C inlet temperature. The carrier gas was helium at 1.1 ml min<sup>-1</sup> constant flow. The temperature gradient was as follows: 2 min at 60 °C, ramp up to 220 °C at 7 °C min<sup>-1</sup>, ramp up to 300 °C at 60 °C min<sup>-1</sup> and 2 min at 300 °C. The MSD transfer line temperature was 280 °C, and the MS source temperature was 230 °C. After a solvent delay of 6 min, a mass range of 35–250 a.m.u. (atomic mass unit) was collected at 70 eV fragmentation energy.

For confirmation of 7*R*-*cis*-*cis* nepetalactone identity, samples were analysed on a chiral column using previously established conditions<sup>23</sup>. Briefly, samples were analysed on an Agilent system consisting of an 8890GC system, a 5977B GC/MSD detector, and a 7693 A autosampler. A Supelco β-DEX 225 column (ID = 0.25 mm; *L* = 30 m; film thickness = 0.25 µm) was used for separation. The samples (1 µl) were injected at 220 °C inlet temperature using a split ratio of 1:10.

Helium was used as carrier gas at a flow rate of 1.1 ml min<sup>-1</sup>. The oven temperature ramp was as follows: 3 min at 80 °C, ramp up to 120 °C at 10 °C min<sup>-1</sup>, 45 min at 120 °C, ramp up to 200 °C at 10 °C min<sup>-1</sup> and 2 min at 200 °C. The MSD transfer line temperature was 220 °C and the MS source at 230 °C. After 10 min of solvent delay, a mass range of 50–350 a.m.u. was collected at 70 eV fragmentation energy.

GC-MS data were collected using Agilent MassHunter Work Station 10.1.49.

### Synthesis of 7*R*-*cis*-*cis* nepetalactone from enzymatically produced 7*R*-*cis*-*cis* nepetalactol



The combined assays (5 × 100 µl reactions with AmISY and AmICYC performed as described above) were extracted with CH<sub>2</sub>Cl<sub>2</sub> (100 µl × 3), and the resulting organic phase was then dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and partially concentrated under a gentle stream of Ar. Then 4 Å molecular sieve (~5 mg), *N*-methylmorpholine *N*-oxide (1 mg, 8.54 µmol) and tetra-*n*-propylammonium perruthenate (1 mg, 2.85 µmol) were sequentially added to the concentrated organic extract. After stirring under an Ar atmosphere at room temperature for 3 h, the reaction mass was filtered through a short pad of Celite and then directly passed through a short silica column, eluting with CH<sub>2</sub>Cl<sub>2</sub> (2 ml) and Et<sub>2</sub>O (2 ml), to give a colourless eluate that was then partially concentrated under a gentle stream of Ar and directly submitted for chiral GC-MS analysis.

### GC-MS data analysis

Data were analysed in Agilent MassHunter Qualitative Analysis 10.0. Total ion chromatograms were exported as CSV files and imported to Prism Graphpad version 10.4.1 for visualization. Total ion chromatograms shown in the same graph are with the same scaled y axis unless otherwise indicated. Peak areas were calculated in Agilent MassHunter Qualitative Analysis 10.0, exported and normalized to internal standard camphor. Bar graphs were constructed in Prism Graphpad version 10.4.1.

### Phylogenetic analyses

Sequences of ICYC orthologues and other MESs were obtained through blast searches against TAIR11, NCBI and One Thousand Plant Transcriptomes (IKP)<sup>91</sup> databases, against transcriptome assemblies from the MintGenomics project<sup>92</sup>, or against genome data obtained during this study. Sequences and accession numbers are provided in Supplementary Dataset 1. Full-length amino acid sequences were aligned with webPRANK (<https://www.ebi.ac.uk/goldman-srv/webprank/>; version updated on 8 October 2017)<sup>93</sup>. Sequence logos were obtained by importing alignments to Geneious Prime 2025.1.2 (Dotmatrix). The IQ-TREE webserver (<http://iqtree.cibiv.univie.ac.at/>) was used to build maximum likelihood phylogenetic trees (automatic substitution model; bootstrap value 1,000) with IQ-Tree 2.3.6 (ref. 94). Trees were visualized in iTOL and graphically edited using itol (<https://itol.embl.de/>)<sup>95</sup> and Adobe Illustrator 27.8.

### Protein models and Docking

Protein models were predicted by AlphaFold3 through the AlphaFold Server (<https://alphafoldserver.com/>)<sup>96</sup>. Docking was done using AutoDock Vina python version 1.2.5 on the SwissDock webserver (<https://www.swissdock.ch/>)<sup>97,98</sup>. Models were visualized using ChimeraX version 1.8 for Mac<sup>99</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All sequencing data associated with this study are available via the NCBI Sequence Read Archive BioProject [PRJNA1270996](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1270996) and [PRJNA1169657](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1169657) (Supplementary Table 1). Reported gene sequences cloned from plant material were deposited to NCBI Genbank under the following accession numbers (Supplementary Table 9): CiGES ([PV988048](https://www.ncbi.nlm.nih.gov/nuccore/PV988048)), CiG8H ([PV988049](https://www.ncbi.nlm.nih.gov/nuccore/PV988049)), Ci8HGO ([PV988050](https://www.ncbi.nlm.nih.gov/nuccore/PV988050)), CiISY ([PV988051](https://www.ncbi.nlm.nih.gov/nuccore/PV988051)), Ci7DLGT ([PV988052](https://www.ncbi.nlm.nih.gov/nuccore/PV988052)), Ci7DLH ([PV988053](https://www.ncbi.nlm.nih.gov/nuccore/PV988053)), CiLAMT ([PV988054](https://www.ncbi.nlm.nih.gov/nuccore/PV988054)), CiSLS ([PV988055](https://www.ncbi.nlm.nih.gov/nuccore/PV988055)), AsGES ([PV988056](https://www.ncbi.nlm.nih.gov/nuccore/PV988056)), AsG8H ([PV988057](https://www.ncbi.nlm.nih.gov/nuccore/PV988057)), As8HGO ([PV988058](https://www.ncbi.nlm.nih.gov/nuccore/PV988058)), AsISY ([PV988059](https://www.ncbi.nlm.nih.gov/nuccore/PV988059)), As7DLGT ([PV988060](https://www.ncbi.nlm.nih.gov/nuccore/PV988060)), As7DLH ([PV988061](https://www.ncbi.nlm.nih.gov/nuccore/PV988061)), AsSLAS ([PV988062](https://www.ncbi.nlm.nih.gov/nuccore/PV988062)), AsICYC ([PV988063](https://www.ncbi.nlm.nih.gov/nuccore/PV988063)), CiICYC ([PV988064](https://www.ncbi.nlm.nih.gov/nuccore/PV988064)) and CrICYC ([PV988065](https://www.ncbi.nlm.nih.gov/nuccore/PV988065)). Source data are provided with this paper.

### References

- Albach, D. C., Soltis, P. S. & Soltis, D. E. Patterns of embryological and biochemical evolution in the asterids. *Syst. Bot.* **26**, 242–262 (2001).
- Stull, G. W., Schori, M., Soltis, D. E. & Soltis, P. S. Character evolution and missing (morphological) data across Asteridae. *Am. J. Bot.* **105**, 470–479 (2018).
- Eisner, T. Catnip: its *raison d'être*. *Science* **146**, 1318–1320 (1964).
- Dobler, S., Petschenka, G. & Pankoke, H. Coping with toxic plant compounds—the insect's perspective on iridoid glycosides and cardenolides. *Phytochemistry* **72**, 1593–1604 (2011).
- Birkett, M. A., Hassanali, A., Hoglund, S., Pettersson, J. & Pickett, J. A. Repellent activity of catmint, *Nepeta cataria*, and iridoid nepetalactone isomers against Afro-tropical mosquitoes, ixodid ticks and red poultry mites. *Phytochemistry* **72**, 109–114 (2011).
- Koudounas, K. et al. A defence-related *Olea europaea* beta-glucosidase hydrolyses and activates oleuropein into a potent protein cross-linking agent. *J. Exp. Bot.* **66**, 2093–2106 (2015).
- Dinda, B., Debnath, S. & Harigaya, Y. Naturally occurring secoiridoids and bioactivity of naturally occurring iridoids and secoiridoids. A review, part 2. *Chem. Pharm. Bull.* **55**, 689–728 (2007).
- Viljoen, A., Mncwangi, N. & Vermaak, I. Anti-inflammatory iridoids of botanical origin. *Curr. Med. Chem.* **19**, 2104–2127 (2012).
- Lichman, B. R. The scaffold-forming steps of plant alkaloid biosynthesis. *Nat. Prod. Rep.* **38**, 103–129 (2021).
- Simkin, A. J. et al. Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry* **85**, 36–43 (2013).
- Collu, G. et al. Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* **508**, 215–220 (2001).
- Miettinen, K. et al. The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 3606 (2014).
- Geu-Flores, F. et al. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* **492**, 138–142 (2012).
- Lichman, B. R. et al. Uncoupled activation and cyclization in catmint reductive terpenoid biosynthesis. *Nat. Chem. Biol.* **15**, 71–79 (2019).
- Kries, H., Kellner, F., Kamileen, M. O. & O'Connor, S. E. Inverted stereocontrol of iridoid synthase in snapdragon. *J. Biol. Chem.* **292**, 14659–14667 (2017).
- Murata, J., Roepke, J., Gordon, H. & De Luca, V. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* **20**, 524–542 (2008).

17. Irmeler, S. et al. Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **24**, 797–804 (2008).
18. Awadasseid, A. et al. Characterization of *Camptotheca acuminata* 10-hydroxygeraniol oxidoreductase and iridoid synthase and their application in biological preparation of nepetalactol in *Escherichia coli* featuring NADP(+) - NADPH cofactors recycling. *Int. J. Biol. Macromol.* **162**, 1076–1085 (2020).
19. Miller, J. C. & Schuler, M. A. Single mutations toggle the substrate selectivity of multifunctional *Camptotheca* secologanic acid synthases. *J. Biol. Chem.* **298**, 102237 (2022).
20. Jensen, S. R. Systematic implications of the distribution of iridoids and other chemical compounds in the Loganiaceae and other families of the Asteridae. *Ann. Missouri Botanical Gard.* **79**, 284–302 (1992).
21. Jensen, S.R. Plant iridoids, their biosynthesis and distribution in angiosperms. In *Annual Proceedings of the Phytochemical Society of Europe. Ecological Chemistry and Biochemistry of Plant Terpenoids*, Vol. 31 (eds Harborne, J. B. & Tomas-Barbaren, F. A.) 133–158 (Clarendon Press/Oxford University Press, 1991).
22. Dudley, Q. M. et al. Reconstitution of monoterpene indole alkaloid biosynthesis in genome engineered *Nicotiana benthamiana*. *Commun. Biol.* **5**, 949 (2022).
23. Hernandez Lozada, N. J. et al. Biocatalytic routes to stereo-divergent iridoids. *Nat. Commun.* **13**, 4718 (2022).
24. Lichman, B. R. et al. The evolutionary origins of the cat attractant nepetalactone in catnip. *Sci. Adv.* **6**, eaba0721 (2020).
25. Zhang, J. et al. A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* **609**, 341–347 (2022).
26. Colinas, M., et al. Ipecac alkaloid biosynthesis in two evolutionarily distant plants. *Nat. Chem. Biol.* <https://doi.org/10.1038/s41589-025-01926-z> (2025).
27. Burlat, V., Oudin, A., Courtois, M., Rideau, M. & St-Pierre, B. Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J.* **38**, 131–141 (2004).
28. Li, C. et al. Cell-type-aware regulatory landscapes governing monoterpene indole alkaloid biosynthesis in the medicinal plant *Catharanthus roseus*. *New Phytol.* **245**, 347–362 (2025).
29. Li, C. et al. Single-cell multi-omics in the medicinal plant *Catharanthus roseus*. *Nat. Chem. Biol.* **19**, 1031–1041 (2023).
30. Sun, S. et al. Single-cell RNA sequencing provides a high-resolution roadmap for understanding the multicellular compartmentation of specialized metabolism. *Nat. Plants* **9**, 179–190 (2023).
31. Van Moerkercke, A. et al. The basic helix-loop-helix transcription factor BIS2 is essential for monoterpene indole alkaloid production in the medicinal plant *Catharanthus roseus*. *Plant J.* **88**, 3–12 (2016).
32. Van Moerkercke, A. et al. The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpene indole alkaloid pathway in *Catharanthus roseus*. *Proc. Natl Acad. Sci. USA* **112**, 8130–8135 (2015).
33. Colinas, M. et al. Subfunctionalization of paralog transcription factors contributes to regulation of alkaloid pathway branch choice in *Catharanthus roseus*. *Front Plant Sci.* **12**, 687406 (2021).
34. Lemenager, D. et al. Purification, molecular cloning, and cell-specific gene expression of the alkaloid-accumulation associated protein CrPS in *Catharanthus roseus*. *J. Exp. Bot.* **56**, 1221–1228 (2005).
35. Zuntini, A.R. et al. Phylogenomics and the rise of the angiosperms. *Nature* **629**, 843–850 (2024).
36. Jensen, S. R. & Nielsen, B. J. Iridoid glucosides in *Griselinia*, *Aralidium* and *Toricellia*. *Phytochemistry* **19**, 2685–2688 (1980).
37. Murayama, T. et al. A secoiridoid, 10-cinnamoyloxyoleoside, isolated from the leaves of *Helwingia japonica*. *Nat. Med.* **58**, 42–45 (2004).
38. Lawas, L. M. F., Kamileen, M. O., Buell, C. R., O'Connor, S. E. & Leisner, C. P. Transcriptome-based identification and functional characterization of iridoid synthase involved in monotropein biosynthesis in blueberry. *Plant Direct* **7**, e512 (2023).
39. Kakuda, R., Imai, M., Yaoita, Y., Machida, K. & Kikuchi, M. Secoiridoid glycosides from the flower buds of *Lonicera japonica*. *Phytochemistry* **55**, 879–881 (2000).
40. Fuji, Y. et al. Chemical characterization and biological activity in young sesame leaves (*Sesamum indicum* L.) and changes in iridoid and polyphenol content at different growth stages. *PLoS ONE* **13**, e0194449 (2018).
41. Yang, Y. et al. Inactive methyl indole-3-acetic acid ester can be hydrolyzed and activated by several esterases belonging to the AtMES esterase family of *Arabidopsis*. *Plant Physiol.* **147**, 1034–1045 (2008).
42. Forouhar, F. et al. Structural and biochemical studies identify tobacco SABP2 as a methyl salicylate esterase and implicate it in plant innate immunity. *Proc. Natl Acad. Sci. USA* **102**, 1773–1778 (2005).
43. Vlot, A. C. et al. Identification of likely orthologs of tobacco salicylic acid-binding protein 2 and their role in systemic acquired resistance in *Arabidopsis thaliana*. *Plant J.* **56**, 445–456 (2008).
44. Hong, B. et al. Biosynthesis of strychnine. *Nature* **607**, 617–622 (2022).
45. Dogru, E. et al. The gene encoding polyneuridine aldehyde esterase of monoterpene indole alkaloid biosynthesis in plants is an ortholog of the alpha/beta-hydroxylase super family. *Eur. J. Biochem.* **267**, 1397–1406 (2000).
46. Farrow, S. C. et al. Biosynthesis of an anti-addiction agent from the iboga plant. *J. Am. Chem. Soc.* **141**, 12979–12983 (2019).
47. Trenti, F. et al. Early and late steps of quinine biosynthesis. *Org. Lett.* **23**, 1793–1797 (2021).
48. Volk, J. et al. Two novel methylesterases from *Olea europaea* contribute to the catabolism of oleoside-type secoiridoid esters. *Planta* **250**, 2083–2097 (2019).
49. Rauwerdink, A. & Kazlauskas, R. J. How the same core catalytic machinery catalyzes 17 different reactions: the serine-histidine-aspartate catalytic triad of alpha/beta-hydroxylase fold enzymes. *ACS Catal.* **5**, 6153–6176 (2015).
50. Ollis, D. L. et al. The  $\alpha/\beta$  hydrolase fold. *Protein Eng.* **5**, 197–211 (1992).
51. Mattern-Dogru, E., Ma, X., Hartmann, J., Decker, H. & Stockigt, J. Potential active-site residues in polyneuridine aldehyde esterase, a central enzyme of indole alkaloid biosynthesis, by modelling and site-directed mutagenesis. *Eur. J. Biochem.* **269**, 2889–2896 (2002).
52. Lindner, S., Geu-Flores, F., Brase, S., Sherden, N. H. & O'Connor, S. E. Conversion of substrate analogs suggests a Michael cyclization in iridoid biosynthesis. *Chem. Biol.* **21**, 1452–1456 (2014).
53. Lichman, B. R., O'Connor, S. E. & Kries, H. Biocatalytic strategies towards [4+2] cycloadditions. *Chemistry* **25**, 6864–6877 (2019).
54. Carrier, G. et al. An efficient and rapid protocol for plant nuclear DNA preparation suitable for next generation sequencing methods. *Am. J. Bot.* **98**, e13–e15 (2011).
55. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).

56. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
57. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
58. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
59. Campbell, M. S. et al. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* **164**, 513–524 (2014).
60. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* <https://doi.org/10.1002/0471250953.bi0410s25> (2009).
61. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
62. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
63. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).
64. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
65. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, lqaa108 (2021).
66. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
67. Lamesch, P. et al. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).
68. Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
69. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
70. Li, W. et al. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43**, W580–W584 (2015).
71. Li, C. et al. Nuclei isolation protocol from diverse angiosperm species. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.11.03.515090> (2022).
72. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.05.05.442755> (2021).
73. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
74. Li, C., Deans, N. C. & Buell, C. R. ‘Simple Tidy GeneCoEx’: a gene co-expression analysis workflow powered by tidyverse and graph-based clustering in R. *Plant Genome* **16**, e20323 (2023).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
76. Afgan, E. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).
77. Cardenas, P. D. et al. Pathways to defense metabolites and evading fruit bitterness in genus *Solanum* evolved through 2-oxoglutarate-dependent dioxygenases. *Nat. Commun.* **10**, 5169 (2019).
78. Berrow, N. S. et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).
79. Odum, M. T. et al. DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. *Nucleic Acids Res.* **52**, W215–W220 (2024).
80. Chen, H. et al. Firefly luciferase complementation imaging assay for protein-protein interactions in plants. *Plant Physiol.* **146**, 368–376 (2008).
81. Zhang, Y. et al. A highly efficient agrobacterium-mediated method for transient gene expression and functional studies in multiple plant species. *Plant Commun.* **1**, 100028 (2020).
82. Fernandez-Pozo, N., Rosli, H. G., Martin, G. B. & Mueller, L. A. The SGN VIGS tool: user-friendly software to design virus-induced gene silencing (VIGS) constructs for functional genomics. *Mol. Plant* **8**, 486–488 (2015).
83. Liscombe, D. K. & O’Connor, S. E. A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. *Phytochemistry* **72**, 1969–1977 (2011).
84. Liu, Y., Schiff, M., Marathe, R. & Dinesh-Kumar, S. P. Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. *Plant J.* **30**, 415–429 (2002).
85. Pollier, J., Vanden Bossche, R., Rischer, H. & Goossens, A. Selection and validation of reference genes for transcript normalization in gene expression studies in *Catharanthus roseus*. *Plant Physiol. Biochem.* **83**, 20–25 (2014).
86. Schmid, R. et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* **41**, 447–449 (2023).
87. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
88. Chapelle, J. P. Vogeloside et acide secologanique, glucosides securiridoides d’*Anthocleista vogelii*. *Planta Med.* **29**, 268–274 (1976).
89. Stavrinides, A. et al. Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity. *Nat. Commun.* **7**, 12116 (2016).
90. Hicks, M. A. et al. The evolution of function in strictosidine synthase-like proteins. *Proteins* **79**, 3082–3098 (2011).
91. One Thousand Plant Transcriptomes, I. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
92. Mint Evolutionary Genomics Consortium. Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol. Plant* **11**, 1084–1096 (2018).
93. Löytynoja, A. & Goldman, N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* **11**, 579 (2010).
94. Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
95. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
96. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
97. Bugnon, M. et al. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Res.* **52**, W324–W332 (2024).
98. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).

99. Meng, E. C. et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
100. Chanderbali, A.S. et al. Draft genome assemblies for two species of *Escallonia* (Escalloniales). *BMC Genom. Data* **25**, 1 (2024).
101. Li, M. et al. Genome structure and evolution of *Antirrhinum majus* L. *Nat. Plants* **5**, 174–183 (2019).
102. Li, Y. et al. High-quality de novo assembly of the *Eucommia ulmoides* haploid genome provides new insights into evolution and rubber biosynthesis. *Hortic. Res.* **7**, 183 (2020).
103. Yocca, A. E. et al. Blueberry and cranberry pangenomes as a resource for future genetic studies and breeding efforts. *Hortic. Res.* **10**, uhad202 (2023).
104. Yu, H. et al. Chromosome-scale genome assembly of an important medicinal plant honeysuckle. *Sci. Data* **9**, 226 (2022).

## Acknowledgements

We thank the gardeners E. Rothe and E. Goschala for growing and maintaining *A. salviifolium* and *C. ipecacuanha* plants as well as F. Kaltofen for growing *N. benthamiana* plants. We thank B. Lichman for providing pOPINF\_NmMLPL construct, Q. M. Dudley for providing NmMLPL and GPPS constructs, and P. Sonawane for providing the modified 3Q1 vector. We are grateful to N. J. Hernández Lozada for providing 7R-cis-cis nepetalactone standard and C. E. Rodríguez-López for providing S-8-oxocitronellal and R-8-oxocitronellal. We thank H. Leucke for assistance with RNA extraction, M. Kunert for assistance with UPLC-MS/MS and GC-MS analyses and Y. Nakamura for NMR structural confirmation of commercial 7S-cis-trans nepetalactol standard. We thank L. Carlton for assistance with nanopore cDNA sequencing. We thank B. Vaillancourt for assistance with sequencing data management. We are grateful to the Max Planck Society for funding and Horizon 2020 (MIAMi, grant number 814645) and the Leibniz Prize, Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 505457618 awarded to S.E.O. We acknowledge funding from the University of Georgia, Georgia Research Alliance and Georgia Seed Development to C.R.B.

## Author contributions

M.C. designed all experimental work, analysed data and supervised the work of C.T., C.M., A.D. and J.W. R.P.D. performed RNA extraction, Illumina sequencing genome sequencing and initial genome assemblies. J.P.H. performed genome annotation. M.C. processed tissue-specific gene expression data and performed co-expression analyses. J.C.W. performed snRNA-seq and data processing. K.G. performed VIGS and qPCRs assisted by L.C. R.M.A. performed chemical oxidation of 7-R-cis-cis nepetalactol enzymatic product. C.M. performed split-Luciferase assays. C.T., J.W. and A.D. performed cloning. J.W. and A.D. performed recombinant protein purifications. M.C., C.T. and J.W. performed reconstitution in *N. benthamiana*.

M.C. performed in vitro enzyme assays. S.H. developed the UPLC-MS/MS method. G.R.T. developed the GC-MS method. A.A.L. provided *C. ipecacuanha* living specimens. M.C., C.R.B. and S.E.O. designed the study. M.C. and S.E.O. wrote the paper with input from all other authors.

## Funding

Open access funding provided by Max Planck Society.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41477-025-02122-6>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41477-025-02122-6>.

**Correspondence and requests for materials** should be addressed to Maite Colinas, C. Robin Buell or Sarah E. O'Connor.

**Peer review information** *Nature Plants* thanks Danijela Mišić, Philipp Zerbe and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

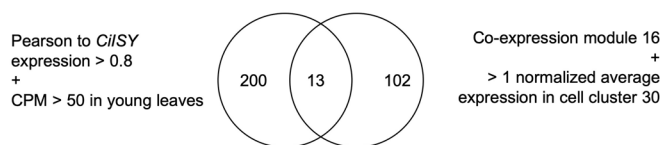
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

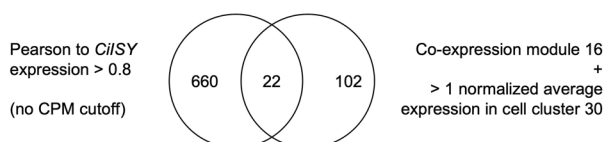
a



## List of 13 genes

| GeneID               | functional annotation                                | name / description | Expression in cell cluster 30 |
|----------------------|--|--------------------|-------------------------------|
| Caipe.S388690        | alcohol dehydrogenase                                | Ci8HGO             | 4.63                          |
| Caipe.S160820        | cytochrome P450, family 76, subfamily C, polypeptide | CiIO               | 4.27                          |
| <b>Caipe.S125200</b> | <b>methyl esterase</b>                               | <b>CiICYC</b>      | <b>3.99</b>                   |
| Caipe.S125190        | cytochrome P450, family 76, subfamily C, polypeptide | CiG8O              | 3.71                          |
| Caipe.S184100        | terpene synthase                                     | CiGES              | 3.64                          |
| Caipe.S366190        | GroES-like zinc-binding dehydrogenase family protein | CYPADH ortholog    | 3.45                          |
| Caipe.S200000        | cytochrome P450, family 72, subfamily A, polypeptide | Ci7DLH             | 3.39                          |
| Caipe.S317530        | NAD(P)-binding Rossmann-fold superfamily protein     | CiISY paralog      | 3.33                          |
| Caipe.S244560        | 1-deoxy-D-xylulose 5-phosphate reductoisomerase      | DXR ortholog       | 3.33                          |
| Caipe.S232250        | Deoxyxylulose-5-phosphate synthase                   | DXS ortholog       | 3.26                          |
| Caipe.S317540        | NAD(P)-binding Rossmann-fold superfamily protein     | CiISY              | 3.04                          |
| Caipe.S364920        | UDP-glucosyl transferase 85A3                        | Ci7DLGT            | 2.81                          |
| Caipe.S351880        | sulfotransferase                                     |                    | 2.57                          |

b

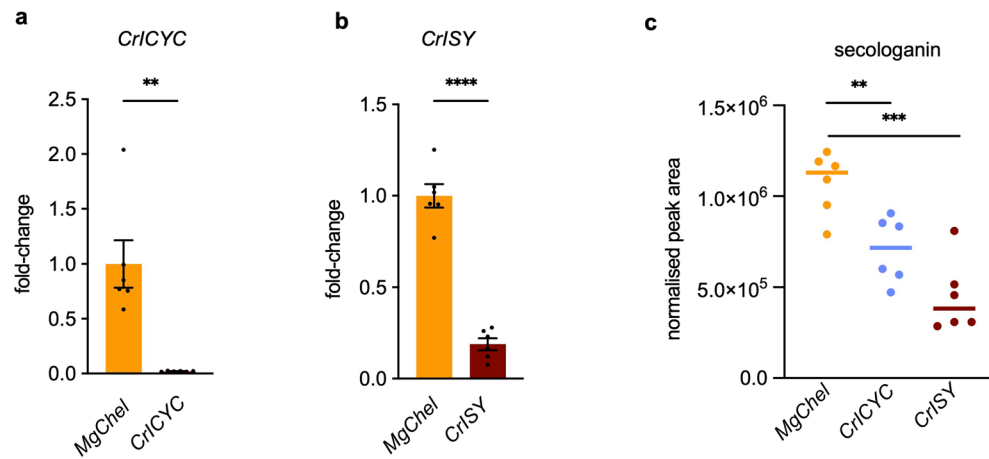


## List of 9 additional genes

| GeneID        | functional annotation   | name / description | Expression in cell cluster 30 |
|---------------|---|--------------------|-------------------------------|
| Caipe.S317500 | F-BOX WITH WD-40  |                    | 3.81                          |
| Caipe.S135190 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein | BIS ortholog       | 2.68                          |
| Caipe.S429940 | Phosphatidylinositol-4-phosphate 5-kinase family protein      |                    | 2.37                          |
| Caipe.S215190 | Acyl-CoA N-acyltransferases (NAT) superfamily protein         |                    | 2.11                          |
| Caipe.S092330 | hypothetical protein  |                    | 2.07                          |
| Caipe.S135260 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein | BIS ortholog       | 1.82                          |
| Caipe.S108310 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein |                    | 1.42                          |
| Caipe.S135200 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein | BIS ortholog       | 1.30                          |
| Caipe.S352430 | Stress responsive alpha-beta barrel domain protein            |                    | 1.11                          |

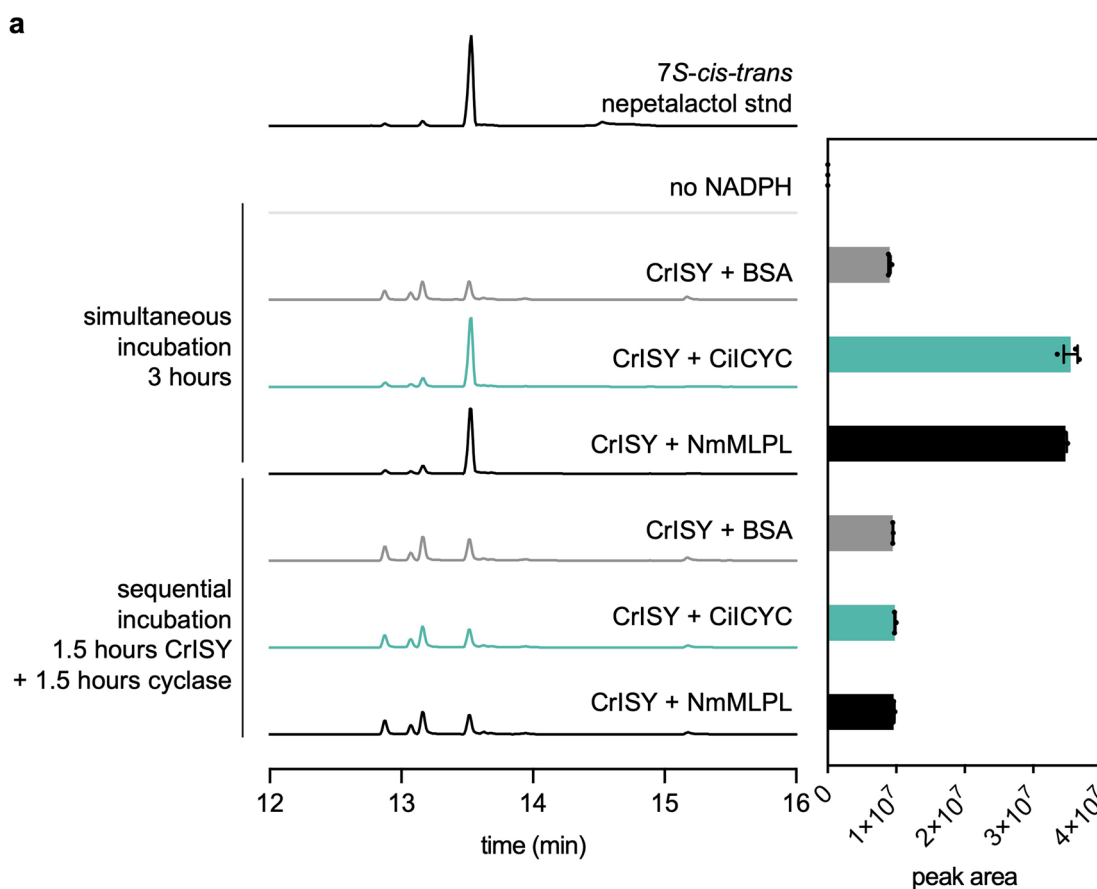
**Extended Data Fig. 1 | Combination of bulk and single nuclei RNA-seq co-expression analysis.** Because it was unclear which type of enzyme would catalyze this cyclization reaction, we attempted to narrow down the list of candidates through co-expression analyses rather than using functional annotation. To achieve the most highly resolved co-expression list, candidate genes from tissue specific co-expression analysis (Pearson correlation to *Ci/SY* expression > 0.8) and cell type specific co-expression analysis (co-expression module 16) were overlaid. **a**, In addition to co-expression, only genes with high absolute expression in bulk RNA-seq (>50 CPM in young leaves from plant 1) and snRNA-seq (normalized average expression > 1 in cell cluster 30) were included. The overlay contained the 13 genes listed with functional annotations based on sequence homology. **ICYC** is shown in bold. Blue indicates orthologs of known

iridoid pathway genes, green indicates orthologs of known MEP pathway genes. CYPADH has been previously associated with iridoid biosynthesis but its exact function could not be determined (Brown et al, 2015). **b**, Without a cutoff for absolute expression in bulk RNA-seq, the list of genes common to both datasets contained 9 additional genes, including orthologs of the *C. roseus* bHLH iridoid synthesis (BIS) transcription factors that are known to induce expression of IPAP specific iridoid pathway genes<sup>31-33</sup>. Blue indicates orthologs of *C. roseus* BIS. CPM, counts per million; GES, geraniol synthase; G8H, geraniol 8-hydroxylase; 8HGO, 8-hydroxygeraniol oxidase; ISY, iridoid synthase; ICYC, iridoid cyclase; IO, iridoid oxidase; 7DLGT, 7-deoxyloganetic acid glucosidase; 7DLH, 7-deoxyloganetic acid hydroxylase.

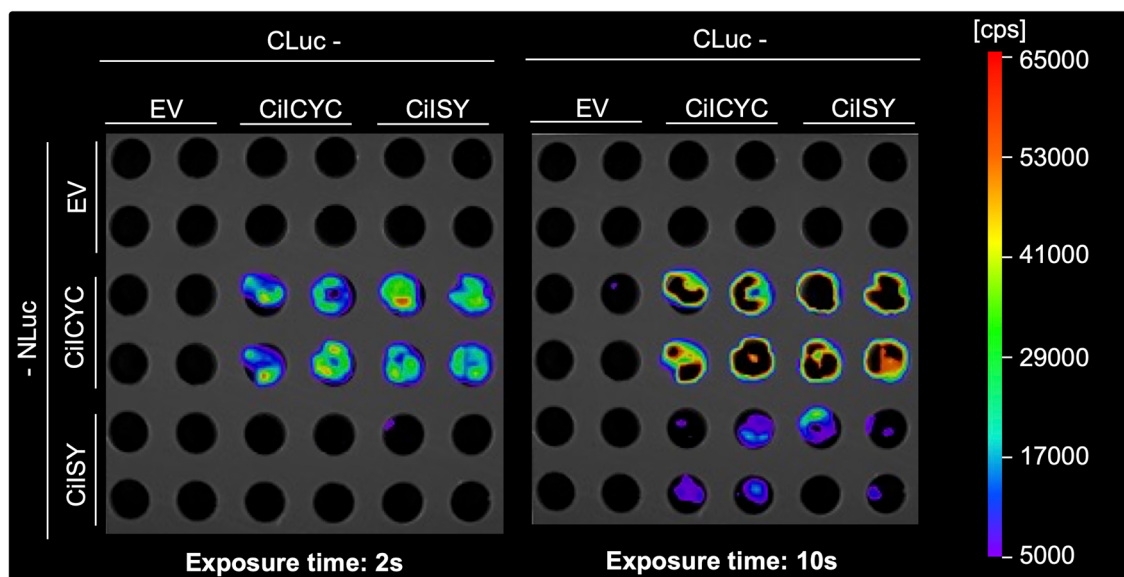


**Extended Data Fig. 2 | Virus Induced Gene Silencing (VIGS) of *ICYC* and *ISY* in *C. roseus*.** Magnesium Chelatase (*MgChel*) was co-silenced in all cases to visualize silenced tissues. In the negative control (in orange) only *MgChel* was silenced. As a positive control *CrISY* was silenced. **a- b**, qPCR confirming the downregulation *CrICYC* (**a**) or *CrISY* (**b**) compared to *MgChel* negative control. Expression values are shown as fold-changes relative to expression in *MgChel* negative control. Bar graphs depict mean N = 6 biological replicates, error bars are standard

error of the mean. Black dot symbols depict values for individual replicates. P-value of unpaired two-tailed ttest with Welch's test correction: \*\*P = 0.0061, \*\*\*\*P < 0.0001. **c**, Secologanin levels in silenced plants compared to control. LC-MS peak areas normalized to internal standard are shown as individual dots of each replicate, the line depicts the mean of N = 6 biological replicates. P-value of unpaired two-tailed ttest with Welch's test correction: \*\*P = 0.0047, \*\*\*P = 0.0002.

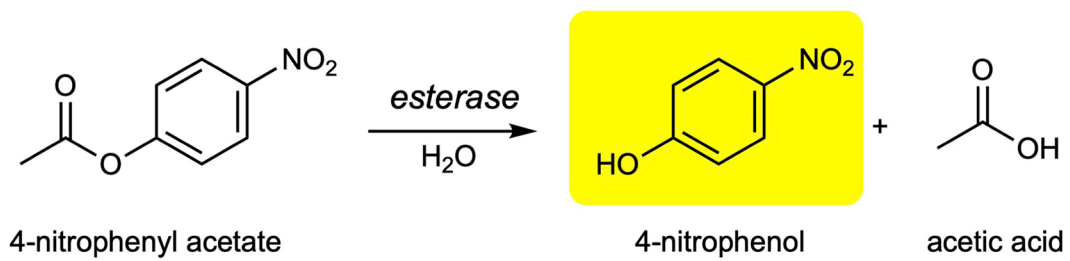
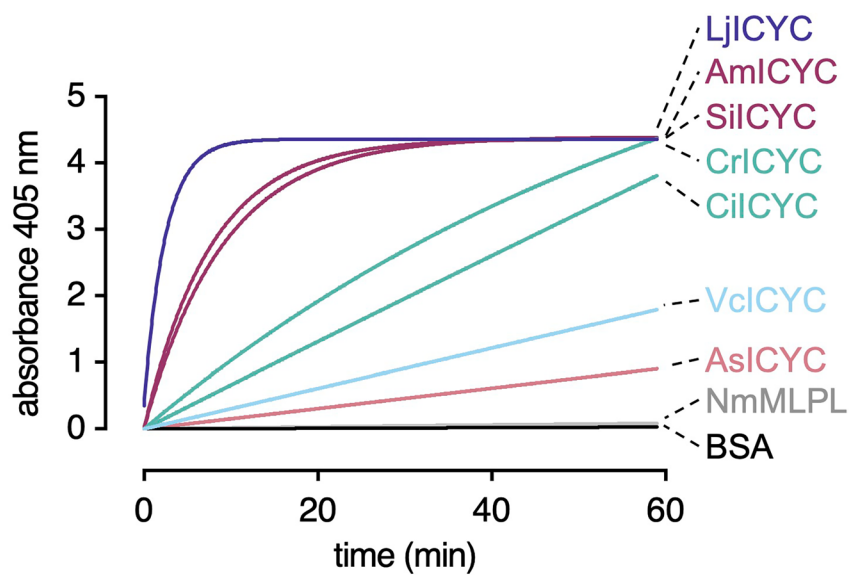
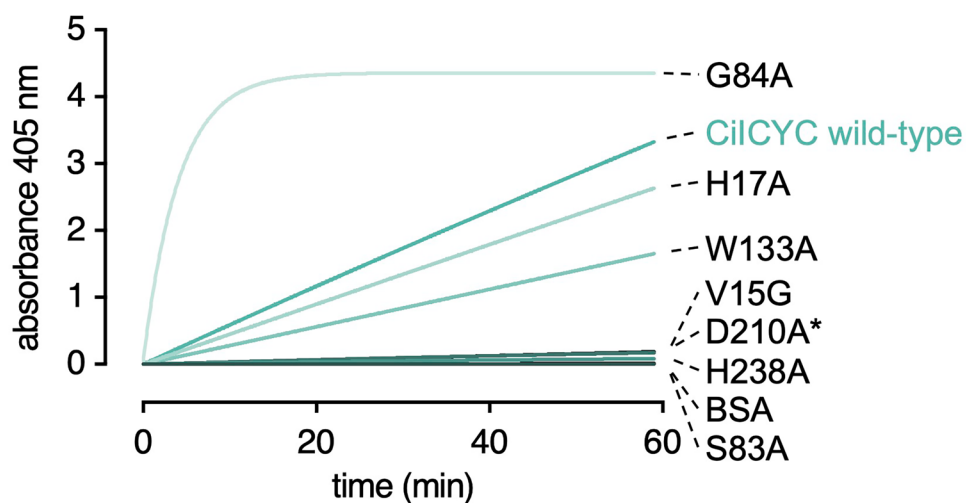


**b**



**Extended Data Fig. 3 | ISY and ICYC must be simultaneously present for nepetalactol formation and may interact.** **a**, Enzyme assays under the conditions shown in Fig. 2a (main text) but with ICYC, NmMLPL or BSA added simultaneously for 3 hours or after 1.5 hours reaction with CrISY alone (sequential incubation). Interestingly, nepetalactol is only formed to higher amounts in simultaneous incubations. This could indicate that the formed ISY product 8-oxocitronellol enol, which is the substrate for the cyclases, is unstable and must be immediately taken up by the cyclases. Bar graphs depict peak areas from  $N = 3$  replicates of nepetalactol, error bars are standard error of the mean. **b**, Split-Luciferase assay with CiICYC and CiISY in *N. benthamiana*. The C-terminal

part of luciferase (CLuc) and the N-terminal part of luciferase (NLuc) are always fused N-terminally or C-terminally, respectively, to the target protein<sup>80</sup>. Empty vectors (EV) contained non-fused CLuc or NLuc, respectively, and served as negative controls. Leaf disks were cut from  $N = 4$  biological replicates of agroinfiltrated *N. benthamiana* plants, transferred to a well plate, where luciferin was added, and imaged using a Nightshade camera (see methods for details). Images are pictures of the same leaf disks with different indicated exposure times. The assays indicate that CiICYC interacts with itself and that CiICYC interacts with CiISY. Note that when NLuc is fused to CiISY interaction is weaker and only revealed at longer exposure times.

**a****b****c**

Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | Esterase activity of ICYC.** **a**, Reaction scheme of the assay. Esterase activity of the proteins was assessed with 4-nitrophenyl acetate as a substrate. The formation of product is assessed by measuring absorbance of the 4-nitrophenol (yellow) at 405 nm. **b**, Esterase activity assays of 0.5  $\mu$ M ICYC orthologs compared to NmMLPL and the negative control BSA. Esterase activity greatly varies among ICYC proteins but is detectable for all ICYC orthologs. **c**, Esterase activity assays of 0.5  $\mu$ M CiCYC wild-type and mutant proteins

(see Fig. 3, main text). Mutations in the catalytic triad led to complete abolishment of esterase activity (the asterisk indicates that D210A mutant was poorly soluble, thus activity could not be assessed reliably). Esterase activity appeared to be poorly correlated with cyclase activity (Fig. 3, main text). Each curve was created by curve fitting of absorbance values of N = 3 replicates. Absorbance was measured every minute on a microplate reader (see methods for details).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

All presented data have been acquired using existing and routinely used software. LC-MS data was collected with Bruker Compass qtofControl 5.2.109 / Hystar 5.1.5.1 or Compass qtofControl 6.3 / Hystar 6.0.30.0 software. GC-MS data was collected using Agilent MassHunter Work Station 10.1.49. Short read RNA-seq and DNA-seq data and single nuclei RNA-seq were sequenced on an Illumina NovaSeq 6000 PE150 platform. Long read DNA-seq was sequenced on a on an R10.4.1 PromethION flowcell (FLO-PRO114M; Oxford Nanopore Technologies (ONT), Oxford, UK). Long read cDNA was sequenced on a MIN106 Rev. D flowcell (Oxford Nanopore Technologies (ONT), Oxford, UK). Split Luciferase data was collected with indigOTM 1.4 software (Berthold Technologies).

## Data analysis

All data analysis was done using routinely used software. For phylogenetic trees sequences were aligned with webPRANK (<https://www.ebi.ac.uk/goldman-srv/webprank/>; version updated 8 Oct 2017). Phylogenetic trees were constructed using IQ-Tree 2.3.6 (<http://iqtree.cibiv.univie.ac.at/>). Protein structural models were predicted using AlphaFold 3 (<https://alphafoldserver.com>). Docking was done using AutoDock Vina python version 1.2.5 on the SwissDock webserver (<https://www.swissdock.ch>). Structures were visualized using ChimeraX 1.8. LC-MS data was analysed using MZmine 4.5.37. GC-MS data was analyzed using Agilent MassHunter Qualitative Analysis 10.0. Chemical structures were generated in ChemDraw Professional 23.1.12. Short read RNA-seq data was quality checked with FastQC (Galaxy Version 0.73), quality trimmed with Trimmomatic (Galaxy Version 0.38.1). Oxford Nanopore (ONT) cDNA reads were processed with Pypochopper v2.7.10. ONT genomic reads were assembled using Flye v.2.9.1. For genome annotation genome assembly data was processed and analyzed using RepeatModeler 2.03, ProtExcuder 1.2, Cutadapt 2.10, HISAT2 2.1.0, minimap2 2.17-r941, Stringtie 2.2.1, BRAKER2 2.1.6, PASA2 2.5.2. Functional annotation of transcripts was performed using BLASTP 2.12.0 and PfamScan 1.6. RNA-seq expression values were generated in CLC Genomics Workbench 24.0.1 (Qiagen). Single nuclei transcriptomics data was processed with pipeseeker 3.1.3, STARsolo 2.7.10b, Seurat 4.3.0.1. Co-expression analysis was done with Simple Tidy\_GeneCoEx. All graphs and heatmaps were prepared with GraphPad Prism 10.4.1 or 10.4.2. Main figures were assembled in Adobe Illustrator 27.8. Supplementary figures were assembled in PowerPoint 16.89.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All sequencing data associated with this study are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive BioProject PRJNA1270996 and PRJNA1169657 (Supplementary Table 1). Reported gene sequences cloned from plant material were deposited to NCBI Genbank under the following accession numbers (Supplementary Table 9): CiGES (PV988048), CiG8H (PV988049), Ci8HGO (PV988050), CiSY (PV988051), Ci7DLGT (PV988052), Ci7DLH (PV988053), CiLAMT (PV988054), CiSLS (PV988055), AsGES (PV988056), AsG8H (PV988057), As8HGO (PV988058), AsISY (PV988059), As7DLGT (PV988060), As7DLH (PV988061), AsSLAS (PV988062), AsICYC (PV988063), CiICYC (PV988064), and CrICYC (PV988065).

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

|  |   |
|--|---|
| Reporting on sex and gender  | <input type="text" value="Not applicable"/> |
| Reporting on race, ethnicity, or other socially relevant groupings | <input type="text" value="Not applicable"/> |
| Population characteristics   | <input type="text" value="Not applicable"/> |
| Recruitment  | <input type="text" value="Not applicable"/> |
| Ethics oversight   | <input type="text" value="Not applicable"/> |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences     Behavioural & social sciences     Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

|                 |   |
|-----------------|---|
| Sample size     | <input type="text" value="Prior determination of sample size is not applicable to this study. Nicotiana benthamiana pathway reconstitution experiments were done on three independent biological replicates which is standard for this type of experiments. In vitro assays were performed as three technical replicates from the same enzyme purifications which is standard for this type of experiments. Bulk RNA-seq was performed on one replicate which is sufficient for gene discovery as no statistics were applied. Single-nuclei RNA-seq was performed on two biological replicates (i.e. individual plants) of Carapichea ipecacuanha which is standard for this type of analysis."/> |
| Data exclusions | <input type="text" value="No data was excluded from the analyses."/>  |

|               |  |
|---------------|--|
| Replication   | Details about biological replicates are provided in the figure legends. Pathway reconstitution experiments were conducted on three biological replicates that correspond to three independent individual plants. All attempts of replication were successful. In vitro assays were performed as three technical replicates from the same enzyme purifications. Replications with different enzyme purifications were successful. |
| Randomization | The order of all LC-MS samples was randomized prior to the runs. Plants were grown in randomized order. Experiments were successfully repeated on other days. For other types of experiments of this study randomization is not relevant.  |
| Blinding      | Blinding was not relevant for this study; characterization of pathway genes and enzymes requires insight into the experimental conditions and characteristics of the samples.  |

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a                                 | Involvement in the study                               |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Plants             |

### Methods

| n/a                                 | Involvement in the study                        |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Dual use research of concern

Policy information about [dual use research of concern](#)

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

| No                                  | Yes   |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Public health              |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> National security          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Crops and/or livestock     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Ecosystems                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other significant area |

### Experiments of concern

Does the work involve any of these experiments of concern:

| No                                  | Yes  |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Demonstrate how to render a vaccine ineffective                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent        |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Increase transmissibility of a pathogen                                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Alter the host range of a pathogen  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable evasion of diagnostic/detection modalities                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable the weaponization of a biological agent or toxin                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other potentially harmful combination of experiments and agents         |

## Plants

|                       |   |
|-----------------------|---|
| Seed stocks           | Nicotiana benthamiana and Catharanthus roseus seeds were obtained from seed stocks maintained by the greenhouse team at Max Planck Institute for Chemical Ecology, Jena. Carapichea ipecacuanha and Alangium salviifolium were obtained as plantlets, no seed stocks were obtained. No other plant species were grown for this study.   |
| Novel plant genotypes | No stable transformation was carried out. Heterologous overexpression in N. benthamiana was done transiently through leaf agroinfiltration as described in the methods paragraph "A. tumefaciens mediated transient expression in N. benthamiana". Virus-induced gene silencing in Catharanthus roseus was also transient as described in the methods paragraph "Virus induced gene silencing (VIGS) in C. roseus". |
| Authentication        | Transient transformation of N. benthamiana through leaf agroinfiltration was done as described in the methods paragraph "A. tumefaciens mediated transient expression in N. benthamiana". Virus-induced gene silencing in Catharanthus roseus was also transient as described in the methods paragraph "Virus induced gene silencing (VIGS) in C. roseus". No stable transformation was carried out.                |